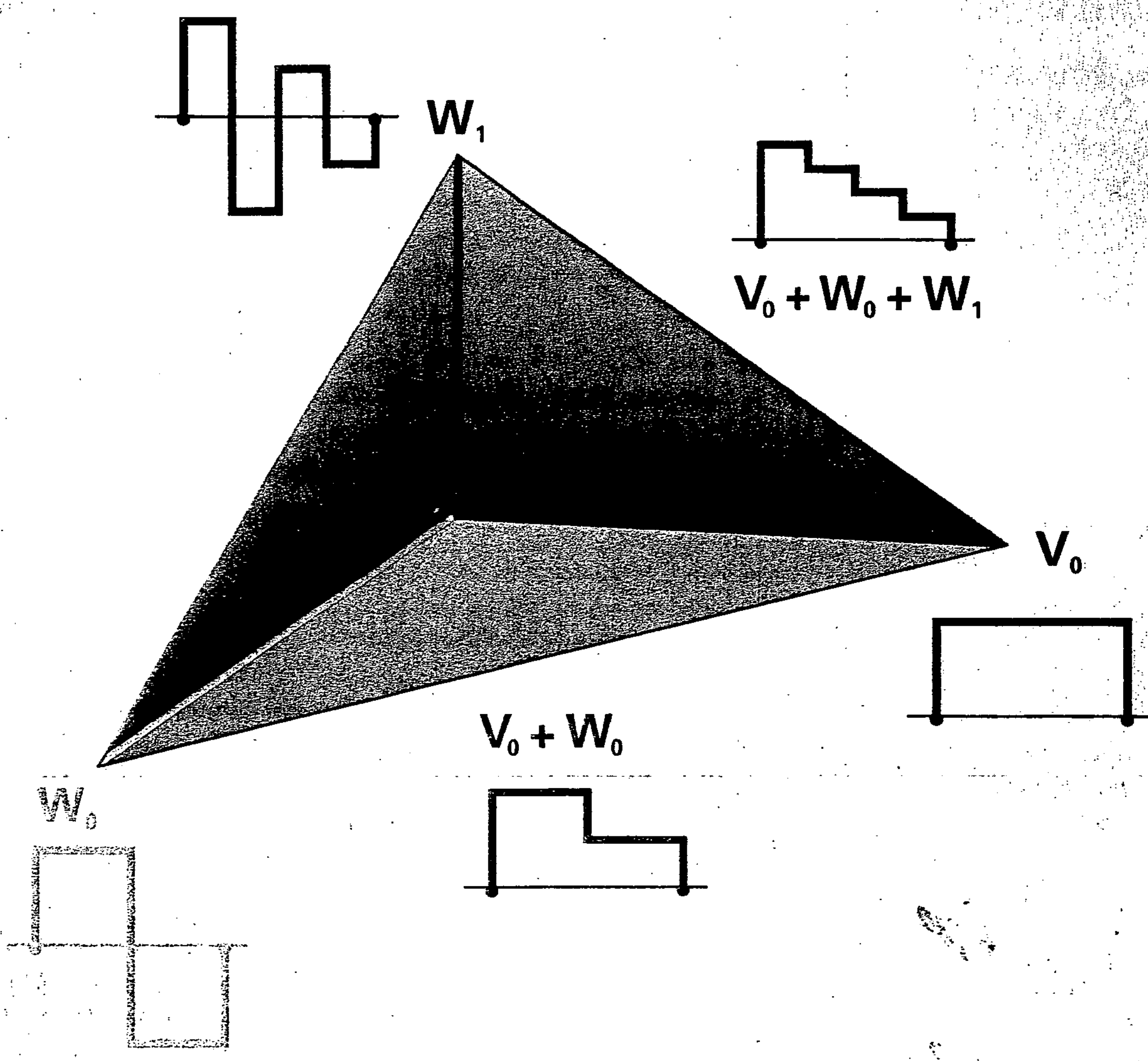
Gilbert Strang / Truong Nguyen

Wavelets and Filter Banks



Wellesley-Cambridge Press

BEST AVAILABLE COPY

Wavelets
and
Filter Banks

Gilbert Strang
Massachusetts Institute of Technology

Truong Nguyen
University of Wisconsin

Wellesley-Cambridge Press
Box 812060
Wellesley MA 02181

Wavelets and Filter Banks

Printed in the United States of America

ISBN 0-9614088-7-1

Other texts from Wellesley-Cambridge Press

Introduction to Applied Mathematics, Gilbert Strang, ISBN 0-9614088-0-4.

An Analysis of the Finite Element Method, Gilbert Strang and George Fix, ISBN 0-13-032946-0.

Calculus, Gilbert Strang, ISBN 0-9614088-2-0.

Introduction to Linear Algebra, Gilbert Strang, ISBN 0-9614088-5-5.

Wellesley-Cambridge Press
Box 812060
Wellesley MA 02181 USA
(617) 431-8488 FAX (617) 253-4358
`gs@math.mit.edu`
`http://www-math.mit.edu/~gs`

MATLAB® is a registered trademark of The MathWorks, Inc. `info@mathworks.com`
24 Prime Park Way Natick MA 01760 (508) 647-7000 `http://www.mathworks.com`

The MATLAB Wavelet Toolbox was created by Michel Misiti, Yves Misiti, Georges Oppenheim, and Jean-Michel Poggi.

A Solutions Manual for instructors is available from the publisher.

Homepage for the book:
`http://saigon.ece.wisc.edu/~waveweb/Tutorials/book.html`

Book design and typesetting using $\LaTeX 2_\varepsilon$ by
Martin Stock, Cambridge, Mass. `mstock@mit.edu`

The cover design is by Gail Corbett and Tracy Baldwin.

Conte

Chapter 1

Introduction

1.1 Overview and Notation

We begin with an overview of ***filters***, ***filter banks***, and ***wavelets***. We want to indicate, first in rough outline and then in detail, the connections between these three topics. Our immediate purpose is to open up the problem and the language — starting with the filter coefficients $h(n)$. The choice of those coefficients is the crucial decision. Their properties govern all that follows.

Each step is a natural development from the one before:

(1) A ***filter*** is a linear time-invariant operator. It acts on input vectors x. The output vector y is the convolution of x with a fixed vector h. The vector h contains the filter coefficients $h(0), h(1), h(2), \ldots$. Our filters are digital, not analog, so the coefficients $h(n)$ come at discrete times $t = nT$. The sampling period T is assumed to be 1 here. The inputs $x(n)$ and outputs $y(n)$ come at all times $t = 0, \pm 1, \pm 2, \ldots$:

$$y(n) \quad = \quad \sum_k h(k)x(n-k) \quad = \quad \text{convolution } h * x \text{ in the time domain.}$$

One input $x = (\ldots, 0, 1, 0, \ldots)$ has special importance — a unit impulse at time zero. The input has $x(n-k) = 0$ except when $n = k$. The sum in the convolution has only one term, and that term is $h(n)$. This output $y(n) = h(n)$ is the response at time n to the unit impulse $x(0) = 1$. It is the *impulse response* $h(0),\ h(1),\ ,\ldots,\ h(N)$.

In a moment the same filter will be described in the frequency domain. Convolution with the vector h will become *multiplication* by a function H. It is the simplicity of multiplication that makes this subject a success. The action of a filter in time and frequency is the foundation on which signal processing is built.

(2) A ***filter bank*** is a set of filters. The analysis bank often has two filters, lowpass and highpass. They separate the input signal into frequency bands. Those subsignals can be compressed much more efficiently than the original signal. Then they can be transmitted or stored. We are describing "subband coding" and its applications. At any time the signals can be recombined (by the *synthesis bank*).

It is not necessary to preserve the full outputs from the analysis filters. Normally they are *downsampled*. ***We keep only the even components of the lowpass and highpass filter outputs.***

If there are M filters, then keeping every Mth component of each output gives a total of the sam length as the input. Critical sampling is the key to subband coding.

This book explains how two or more filters, with downsampling, can jointly achieve proper ties that are impossible for a single filter. We are particularly interested in "perfect reconstructio FIR filter banks". In this case the reconstructed output $\widehat{x}(n)$ from the synthesis bank is identica to the original input x to the analysis bank (with only a time delay). In matrix language, a bande matrix (for the analysis bank) has a banded inverse (the synthesis bank).

In the frequency domain, each filter leads to a multiplication. But downsampling is *not a time-invariant operation*. If we delay all components of y by one time unit, the output from downsampling is totally different. The new samples $y(-1), y(1), y(3)$ are entirely separate and independent from the original samples $y(0), y(2), y(4)$. Those two subsampled signals are two "phases" of y, not connected. Therefore downsampling alters the multiplication picture in the frequency domain. In fact it introduces *aliasing*.

Chapter 4 will show how the simplicity of multiplication can be rescued by looking at each phase separately. Each phase of y comes from filtering the phases of x (using phases of h). These separate pieces are multiplications in the frequency domain. The whole operation together, filtering followed by downsampling, becomes a matrix multiplication — by the *polyphase matrix.* This is the foundation of filter bank theory (still to be explained in detail!). The analysis polyphase matrix H_p will reveal the correct synthesis bank for perfect reconstruction. That synthesis filter bank uses H_p^{-1}.

(3) *Wavelets* are basis functions $w_{jk}(t)$ in continuous time. A basis is a set of linearly independent functions that can be used to produce all admissible functions $f(t)$:

$$\boxed{f(t) = \text{ combination of basis functions } = \sum_{j,k} b_{jk}\, w_{jk}(t).} \tag{1.1}$$

The special feature of the wavelet basis is that all functions $w_{jk}(t)$ are constructed from a single mother wavelet $w(t)$. This wavelet is a small wave (a pulse). Normally it starts at time $t = 0$ and ends at time $t = N$.

The shifted wavelets w_{0k} start at time $t = k$ and end at time $t = k+N$. The rescaled wavelets w_{j0} start at time $t = 0$ and end at time $t = N/2^j$. Their graphs are compressed by the factor 2^j, where the graphs of w_{0k} are translated (shifted to the right) by k:

$$\textit{compressed:}\quad w_{j0} = w\left(2^j t\right) \qquad\qquad \textit{shifted:}\quad w_{0k}(t) = w(t-k).$$

A typical wavelet w_{jk} is compressed j times and shifted k times. Its formula is

$$\boxed{w_{jk}(t) = w\left(2^j t - k\right).}$$

The remarkable property that is achieved by many wavelets is *orthogonality*. The wavelets are orthogonal when their "inner products" are zero:

$$\int_{-\infty}^{\infty} w_{jk}(t)\, w_{JK}(t)\, dt = \text{ inner product of } w_{jk} \text{ and } w_{JK} = 0. \tag{1.2}$$

In this case the wavelets form an *orthogonal basis* for the space of admissible functions. This basis corresponds to a set of axes that meet at 90° angles — as most good axes do. Orthogonality leads to a simple formula for each coefficient b_{JK} in the expansion for $f(t)$. Multiply the

expansion displayed in equation (1.1) by $w_{JK}(t)$ and integrate:

$$\int_{-\infty}^{\infty} f(t)\, w_{JK}(t)\, dt = b_{JK} \int_{-\infty}^{\infty} (w_{JK}(t))^2\, dt. \tag{1.3}$$

All other terms in the sum disappear because of orthogonality. Equation (1.2) eliminates all integrals of w_{jk} times w_{JK}, except the one term that has $j = J$ and $k = K$. That term produces $(w_{JK}(t))^2$. Then b_{JK} is the ratio of the two integrals in equation (1.3).

As we describe the connection between filter banks and wavelets, you will see that it is the "*highpass filter*" that leads to $w(t)$. The "*lowpass filter*" leads to a scaling function $\phi(t)$. In most constructions the lowpass filter comes first—***the scaling function is obtained before the wavelet.*** In fact the scaling function (in continuous time) comes from infinite repetition $L\,L\,\ldots\,L$ of the lowpass filter, with rescaling at each iteration. The wavelet follows from $\phi(t)$ by just *one* application of the highpass filter.

Multiresolution

At a given resolution of a signal or an image, the scaling functions $\phi\left(2^j t - k\right)$ are a basis for the set of signals. The level is set by j, and the time steps at that level are 2^{-j}. The new details at level j are represented by the wavelets $w\left(2^j t - k\right)$. Then the smooth signal plus the details, the ϕ's plus the w's, combine into a ***multiresolution*** of the signal at the finer level $j+1$. Averages come from the scaling functions, details come from the wavelets:

$$\begin{array}{ccc} \text{signal at level } j \text{ (local averages)} & \searrow & \\ + & & \text{signal at level } j+1 \\ \text{details at level } j \text{ (local differences)} & \nearrow & \end{array}$$

That is multiresolution for one signal. When we apply it to all signals, we have multiresolution for *spaces* of functions:

$$\begin{array}{ccc} V_j = \text{scaling space at level } j & \searrow & \\ \oplus & & V_{j+1} = \text{scaling space at level } j+1 \\ W_j = \text{wavelet space at level } j & \nearrow & \end{array}$$

This idea of multiresolution is absolutely basic to wavelet analysis. Again, we are only introducing it. We are sending a coarse signal to the reader, not the details. You only have the input at level 1.

Thus the signal is divided into different ***scales*** of resolution, rather than different frequencies. The "time-scale plane" takes the place for wavelets that the "time-frequency plane" takes for filters. Multiresolution divides the frequencies into *octave bands*, from ω to 2ω, instead of uniform bands from ω to $\omega + \Delta\omega$. The compression of a graph, when $f(t)$ is replaced by $f(2t)$, means expansion of its Fourier transform from $F(\omega)$ to $\frac{1}{2}F\left(\frac{\omega}{2}\right)$. Frequencies shift upward by an octave, when time is rescaled by two. You will see how the time-frequency plane is partitioned naturally into *rectangles of constant area* (Figure 1.1).

This matching of long time with low frequency and short time with high frequency occurs in a natural way for wavelets. It is one of the attractions of a wavelet decomposition.

To the reader: We have reprinted in Appendix A an article on wavelets published in the American Scientist of May 1994. This article introduces wavelet notation through its correspondence with *musical notation*. In music, each note specifies a frequency and a position in time.

Chapter 6

Multiresolution

6.1 The Idea of Multiresolution

Our main approach to wavelets is through 2-channel filter banks. Everything develops from the filter coefficients. All constructions are concrete and highly explicit. Choose good coefficients and you get good wavelets. The heart of the theory is to see how conditions on the numbers $h(k)$ and $c(k)$ and $d(k)$ determine properties of $\phi(t)$ and $w(t)$—the scaling function and the basic wavelet. Then the problem is to design filters that achieve those properties.

By iterating the filter bank, Section 6.2 reaches the ***dilation equation*** for $\phi(t)$ and the ***wavelet equation*** for $w(t)$. Sections 6.3 and 6.4 study those equations in the time domain and frequency domain. Conditions O and A lead to orthogonality and approximation accuracy. The Daubechies wavelets are "optimal" with respect to those two properties. But these orthogonal wavelets are not and cannot be symmetric (except for Haar). Also the transition from passband to stopband is not sharp. So the design problem is still open. Better wavelets remain to be constructed.

This opening section aims for an overview that brings out the key ideas. Before the construction using discrete time, we describe what is wanted in continuous time. The goal is a decomposition of the whole function space into subspaces. That implies a decomposition of each function—*there is a piece of $f(t)$ in each subspace*. Those pieces (or projections) give finer and finer details of $f(t)$. The signal is "resolved" at scales $\Delta t = 1, 1/2, \ldots, (1/2)^j$.

For audio signals, these scales are essentially *octaves*. They represent higher and higher frequencies. For images and indeed for all signals, the simultaneous appearance of multiple scales is known as ***multiresolution***.

Multiresolution will be described first for subspaces V_j and W_j. The scaling spaces V_j are *increasing*. The wavelet space W_j is the *difference* between V_j and V_{j+1}. *The sum of V_j and W_j is V_{j+1}*. Then these extra conditions involving dilation to $2t$ and translation to $t-k$ define a genuine multiresolution:

> If $f(t)$ is in V_j then $f(t)$ and $f(2t)$ and all $f(t-k)$ and $f(2t-k)$ are in V_{j+1}.

In the end, one wavelet generates a whole basis. The functions $w(2^j t - k)$ come by dilation and translation (all j and all k). There are six steps toward this goal, and we take them one at a time:

1. An increasing sequence of subspaces V_j (complete in L^2)

2. The wavelet subspace W_j that gives $V_j + W_j = V_{j+1}$
3. The dilation requirement from $f(t)$ in V_j to $f(2t)$ in V_{j+1}
4. The basis $\phi(t-k)$ for V_0 and $w(t-k)$ for W_0
5. The basis $\phi(2^j t - k)$ for V_j and $w(2^j t - k)$ for W_j
6. The basis of all wavelets $w(2^j t - k)$ for the whole space L^2.

shortcut to multiresolution. Before those six steps, may I mention one shortcut step that ırts with the filter coefficients $h(k)$. That step is *to solve the dilation equation for the scaling ınction* $\phi(t)$:

$$\boxed{\phi(t) = \sum 2\,h(k)\,\phi(2t-k).}$$

ıe first requirements on the coefficients are $\sum h(k) = 1$ and $\sum(-1)^k h(k) = 0$. The full quirement is Condition E in Section 7.2. When this is satisfied, $\phi(t)$ can be computed. Then $\{\phi(2^j t - k)\}$ is a basis for V_j. These spaces are automatically increasing and complete and shift-variant and connected by dilation. Thus multiresolution is achieved.

Scale of Subspaces

ach V_j is contained in the next subspace V_{j+1}. A function in one subspace is in all the higher iner) subspaces:

$$V_0 \subset V_1 \subset \cdots \subset V_j \subset V_{j+1} \subset \cdots$$

. function $f(t)$ in the whole space has a piece in each subspace. Those pieces contain more and ıore of the full information in $f(t)$. The piece in V_j is $f_j(t)$. One requirement on the sequence f subspaces is *completeness*:

$$f_j(t) \to f(t) \qquad \text{as } j \to \infty.$$

The first example will not have the dilation feature required for multiresolution:

:xample 6.1. *V_j contains all trigonometric polynomials of degree $\le j$.*
:ertainly V_j is contained in V_{j+1}. The spaces are growing. (Since Daubechies uses $-j$ where we .se j, her subspaces are *decreasing*. Most authors now use an increasing sequence, for simpler ıumbering.) The piece of $f(t)$ in V_j is the partial sum $f_j(t)$ of its Fourier series:

$$f_j(t) = \sum_{|k| \le j} c_k e^{ikt} \quad \text{is the piece in } V_j.$$

This is the *projection* of $f(t)$ onto V_j. The exponentials e^{ikt} are orthogonal, so the energy in $f_j(t)$ is the sum of $|c_k|^2$ over low frequencies $|k| \le j$. The energy in $f(t) - f_j(t)$ is the sum ›ver high frequencies $|k| > j$. This approaches zero as $j \to \infty$. Therefore the sequence V_j is :omplete in the whole 2π-periodic space L^2.

Now we identify the second family of subspaces. *W_j contains the new information* $\Delta f_j(t) = f_{j+1}(t) - f_j(t)$. This is the "detail" at level j. From the viewpoint of individual functions,

$$\boxed{f_j(t) + \Delta f_j(t) = f_{j+1}(t).} \tag{6.1}$$

From the viewpoint of the subspaces they lie in, this is

$$V_j \bigoplus W_j = V_{j+1}.$$

Each function in V_{j+1} is the sum of two orthogonal parts, f_j in V_j and Δf_j in W_j. In our e ple, the new information $\Delta f_j = f_{j+1} - f_j$ is the new term that enters $f_{j+1}(t)$:

$$\Delta f_j(t) = c_{j+1}e^{i(j+1)t} + c_{-j-1}e^{-i(j+1)t}.$$

The space W_j contains terms of exact degree $j+1$. Those are orthogonal to all terms of de $\leq j$. Together, these orthogonal complements W_j and V_j produce V_{j+1}. This is an impo part of multiresolution.

> *The spaces W_j are **differences** between the V_j.*
> *The spaces V_j are **sums** of the W_j.*

We can call V_j a partial sum by recognizing how the W's add to V. Start from

$$V_0 \bigoplus W_0 = V_1 \quad \text{and} \quad V_1 \bigoplus W_1 = V_2.$$

Substituting the first into the second, V_2 is the sum of three mutually orthogonal subspaces

$$V_0 \bigoplus W_0 \bigoplus W_1 = V_2.$$

When you add details up to and including W_j, you have V_{j+1}:

$$V_0 \bigoplus W_0 \bigoplus W_1 \bigoplus \cdots \bigoplus W_j = V_{j+1}.$$

For the functions in those subspaces, this equation is just

$$f_0(t) + \Delta f_0(t) + \Delta f_1(t) + \cdots + \Delta f_j(t) = f_{j+1}(t).$$

That sum is $f_0 + (f_1 - f_0) + (f_2 - f_1) + \cdots + (f_{j+1} - f_j)$. It "telescopes" into f_{j+1}.

In practice, we can construct the spaces V_j and take differences. Or we can construc W_j and take sums. It is like differentiating an integral or integrating a derivative. Starting a basis is like starting from the W_j; then the V_j are partial sums as above. For the expone basis e^{ikt}, this works fine. It also works for Haar wavelets, which are so simple. But for wavelets, that direct search for $w(t)$ was very difficult.

Important. The construction of wavelets has succeeded by *finding the V_j first.* We begin the scaling function $\phi(t)$, not the wavelet! Its translates $\phi(t-k)$ go into V_0. Rescaling t gives V_j. Then the wavelet spaces W_j are the differences between V_{j+1} and V_j. The func in W_j are the details at the jth scale.

Similarly for filter banks, we design the lowpass filter by choosing $c(k)$. Then the high coefficients $d(k)$ are easy. To maintain this analogy between continuous and discrete, we dr "logarithmic tree" of filter banks. At each step, the highpass filter (with downsampling as u produces the detail Δf_j in W_j. The space V_{j+1} is the sum of $V_0, W_0, \ldots, W_j$.

It is convenient if W_j is orthogonal to V_j. Each W_j is then automatically orthogonal other W_k. *Reason:* If $k < j$, then W_k is contained in V_j — which is perpendicular to W_j. completeness condition can be restated as

$$V_0 \bigoplus \sum_{j=0}^{\infty} W_j = L^2.$$

With orthogonality of each piece $f_j(t)$ to the next detail $\Delta f_j(t)$, these subspaces are orthogonal. But we emphasize now that *orthogonality is not essential.*

A nonorthogonal example comes directly from any nonorthogonal basis $b_0(t)$, $b_1(t)$, ... The piece $f_j(t)$ includes all the terms through $b_j(t)$:

$$\begin{array}{rll} \text{Sum up to } j: & f_j(t) = \sum_0^j c_k b_k(t) & \text{is in } V_j \\ \text{Next term}: & \Delta f_j(t) = c_{j+1} b_{j+1}(t) & \text{is in } W_j. \end{array}$$

The pattern is not lost, just the orthogonality. The new space V_{j+1} is still the "*direct sum*" of V_j and W_j, which intersect only at the zero vector. The angle between subspaces can be less than 90°, as long as every f_{j+1} in V_{j+1} has exactly one splitting into $f_j + \Delta f_j$:

$$\boxed{V_j \cap W_j = \{0\} \quad \text{and} \quad V_j + W_j = V_{j+1}} \tag{6.6}$$

This nonorthogonal situation applies to *biorthogonal* filters and wavelets (Section 6.5). There W_j is orthogonal to a different subspace $\widetilde{V}_j$. The extra freedom can be put to good use.

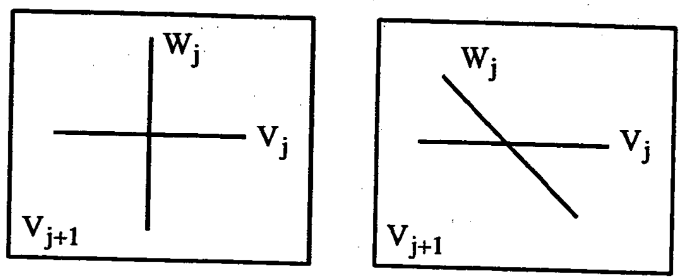


Figure 6.1: An orthogonal sum and a direct sum. Both written $V_j \oplus W_j = V_{j+1}$ and both allowed.

The Dilation Requirement

So far we have an increasing and complete scale of spaces. Each V_j is contained in the next V_{j+1}. For multiresolution, the crucial word *scale* carries an additional meaning. V_{j+1} consists of all rescaled functions in V_j:

$$\textbf{\textit{Dilation:}} \quad f(t) \text{ is in } V_j \quad \Longleftrightarrow \quad f(2t) \text{ is in } V_{j+1}.$$

The graph of $f(2t)$ changes twice as fast as the graph of $f(t)$. On a map, the scale is doubled. At 3,000,000:1 the state of Utah fills a page. At 6,000,000:1 its height is a half page. The length that represents a mile is cut in half. This length is Δt or Δx or h.

The example using $f(t) = c_{-j}e^{-ijt} + \cdots + c_j e^{ijt}$ does not meet this rescaling requirement. The highest frequency only increases by one, between V_j and V_{j+1}. But when t is changed to $2t$, the highest frequency becomes $2j$. *The frequencies must double.* The new space V_{j+1} is required to contain all those new frequencies. To satisfy the scaling requirement, the partial sums go an octave at a time. *The sum for f_j should stop at frequency 2^j instead of j.* Then Δf_j contains all frequencies between 2^j and 2^{j+1}:

$$\begin{array}{rcll} \textbf{\textit{Multiresolution example}}: & f_j(t) & = & \sum c_k e^{ikt} \text{ for } |k| \le 2^j \\ \textbf{\textit{Next detail}}: & \Delta f_j(t) & = & \sum c_k e^{ikt} \text{ for } 2^j < |k| \le 2^{j+1}. \end{array}$$

This is a genuine multiresolution, in which V_j and W_j have roughly the same dimension. It is the Littlewood-Paley decomposition of a Fourier series, into octaves instead of single terms.

This is a chief part of the mathematical background. To fit the requirements precisely, when $f(t)$ is defined on the whole line $-\infty < t < \infty$, we should use all frequencies ω and not just integers:

$$f_j(t) = \frac{1}{2\pi} \int_{|\omega| \le 2^j} \widehat{f}(\omega) e^{i\omega t} d\omega.$$

Now the spaces V_j go down the scale toward $j = -\infty$, as well as up the scale. The continuous frequency ω can be halved as well as doubled. The basis functions become *sinc functions*, by the sampling theorem. *Continuous frequency but discrete basis*, as is normal for L^2. And the nested spaces include $j < 0$:

$$\cdots \subset V_{-1} \subset V_0 \subset \cdots \subset V_j \subset V_{j+1} \subset \cdots \tag{6.7}$$

In addition to completeness as $j \to \infty$, we require emptiness as $j \to -\infty$:

$$\bigcap V_j = \{0\} \quad \text{and} \quad \overline{\bigcup V_j} = \text{ whole space.} \tag{6.8}$$

Emptiness means that $\|f_j(t)\| \to 0$ as $j \to -\infty$. Completeness still means that $f_j(t) \to f(t)$ as $j \to \infty$. The detail $\Delta f_j = f_{j+1} - f_j$ belongs to W_j and we still have

$$V_j \bigoplus W_j = V_{j+1}. \tag{6.9}$$

This can be an orthogonal sum, with Δf_j orthogonal to f_j. It must be a direct sum, with $V_j \bigcap W_j = \{0\}$. The reconstruction of $f(t)$ from its details Δf_j can start at $j = 0$ as before, or it can start at $j = -\infty$:

$$f(t) = f_0(t) + \sum_0^\infty \Delta f_j(t) \quad \text{or} \quad f(t) = \sum_{-\infty}^\infty \Delta f_j(t).$$

The sum of subspaces can start at $j = 0$ or $j = -\infty$. When the sum stops at $J \ge 0$, we have the subspace V_{J+1}:

$$V_{J+1} = V_0 + \sum_{j=0}^J W_j \quad \text{or} \quad V_{J+1} = \sum_{j=-\infty}^J W_j.$$

The left sum includes the scaling functions in V_0. The sum on the right involves only the wavelets. That form includes all the very large time scales $\Delta t = 2^{-j}$ as $j \to -\infty$.

In practice we use the first sum. Our calculations begin at some unit scale. The scaling functions at $j = 0$ and the wavelets with $j \ge 0$ are the basis. I suppose the scaling functions at level $j = J$ and the wavelets with $j \ge J$ are another basis.

The Translation Requirement and the Basis

Instead of rescaling $f(t)$, we now shift its graph. This is *translation*, and it leads to the fundamental requirement of time-invariance in signal processing. The subspaces are *shift-invariant*:

If $f_j(t)$ is in V_j then so are all its translates $f_j(t-k)$.

Suppose $f(t)$ is in V_0. Then $f(2t)$ is in V_1 and so is $f(2t-k)$. By induction, $f(2^jt)$ is in V_j and so is $f(2^jt-k)$. Dilation and translation are now built in.

With translation we are committed to working on the whole line $-\infty < t < \infty$, or to periodicity. A particular $f(t)$ may have compact support, but the whole space V_0 (all functions together) is shift-invariant. For finite intervals, the requirements have to be (and can be) adjusted. Dilation and translation operate freely on the whole line, and can be studied by Fourier transform.

The final requirement for multiresolution concerns a ***basis*** for each space V_j. If we choose one function $\phi(t)$ in V_0, its translates $\phi(t-k)$ may be independent. These translates may span the whole space V_0. They may even be orthonormal. The starting assumption, to be weakened later, is that V_0 contains such a function:

There exists $\phi(t)$ so that $\{\phi(t-k)\}$ is an orthonormal basis for V_0.

When the functions $\phi(t-k)$ are an orthonormal basis for V_0, the rescaled functions $\sqrt{2}\phi(2t-k)$ will be an orthonormal basis for V_1. At scaling level j, the basis functions $\phi(2^jt-k)$ are normalized by $2^{j/2}$. We collect all the requirements in one place:

Multiresolution Analysis

The subspaces V_j satisfy requirements 1 to 4:

1. $V_j \subset V_{j+1}$ and $\bigcap V_j = \{0\}$ and $\overline{\bigcup V_j} = L^2$ (completeness).
2. *Scale invariance*: $f(t) \in V_j \Longleftrightarrow f(2t) \in V_{j+1}$.
3. *Shift invariance*: $f(t) \in V_0 \Longleftrightarrow f(t-k) \in V_0$.
4. *Shift-invariant basis*: V_0 has an orthonormal basis $\{\phi(t-k)\}$.

4$'$. *Shift-invariant basis*: V_0 has a stable basis (Riesz basis) $\{\phi(t-k)\}$.

4 and 4$'$ are interchangeable. A stable basis can be orthogonalized in a shift-invariant way. This is in Section 6.4, together with the definition: stable = Riesz = uniformly independent. In practice we choose a convenient basis, orthogonal or not. Then V_j has the basis $\phi_{jk}(t) = 2^{j/2}\phi(2^jt-k)$:

$$f_j(t) = \sum_{k=-\infty}^{\infty} a_{jk}\phi_{jk}(t) \quad \text{is the piece in } V_j.$$

In the orthogonal case, the energy in this piece is

$$\|f_j\|^2 = \sum_{k=-\infty}^{\infty} |a_{jk}|^2 . \tag{6.10}$$

Shift-invariance and scale-invariance are built in through the basis $\{2^{j/2}\phi(2^jt-k)\}$. This basis combines requirements 2, 3, and 4!

We have at least three ways to construct or describe a multiresolution:

1. By the spaces V_j

2. By the scaling function $\phi(t)$

3. By the coefficients $2h(k)$ in the dilation equation.

Our next examples use the spaces V_j and their bases. Then we move to description 3 and the dilation equation. Section 6.4 will orthogonalize the basis. The result will be the orthonormal $\phi(t-k)$ that multiresolution originally asks for. What we really need is a good shift-invariant basis.

It is also possible to allow *several* scaling functions $\phi_1, \ldots, \phi_r$, when one function (with its translates) cannot produce the whole space V_0. This occurs in Example 3 below. It corresponds to "*multiwavelets*".

The framework for multiresolution is set by the dilation-translation requirement. Examples come first. Then we study the dilation equation, and construct wavelets.

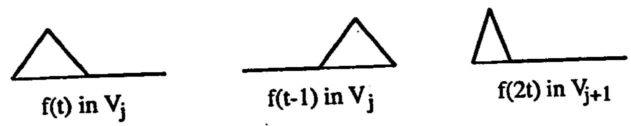


Figure 6.2: Translation stays in V_j. Dilation moves into V_{j+1}. Why is $f(2t) - f(t)$ *not* in W_j?

Examples of Multiresolution

1. Piecewise constant functions. V_0 contains all functions in L^2 that are constant on unit intervals $n \le t < n+1$. These functions are determined by their values $f(n)$ at all integer times $t = n$:

$$f(t) = f(\text{integer part of } t).$$

The function $f(2t)$ in V_1 is then constant on half-intervals. The functions in V_j are constant on intervals of length 2^{-j}. The spaces are increasing, $V_j \subset V_{j+1}$, because any function that is constant on intervals of length 2^{-j} is automatically constant on intervals of half that length. These are *dyadic intervals*, starting at a dyadic number $t = n/2^j$ and ending at $t = (n+1)/2^j$.

These spaces are shift-invariant—the translate of a piecewise constant function is still piecewise constant. The step from j to $j+1$ rescales time by 2 and produces V_{j+1}. What about a basis? The simplest choice is the *box function*:

$$\phi(t) = \begin{cases} 1 & \text{for } 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{is orthogonal to its translates } \phi(t-k).$$

Every function in V_0 is a combination of boxes $f(t) = \sum f(n)\phi(t-n)$. So requirement 4 i satisfied by the box function $\phi(t)$.

2. Continuous piecewise linear functions. The functions $f(t)$ are now linear between eac pair of values $f(n)$ and $f(n+1)$. Notice again the shift-invariance and the scale-invariance:

Shift: If $f(t)$ is piecewise linear, so is $f(t-k)$.
Scale: If $f(t)$ is linear on unit intervals, then $f(2t)$ is linear on half-intervals.

The spaces are right for multiresolution. Is there a shift-invariant basis?

The basis function that comes to mind is the *hat function* $H(t)$, equal to one at $t = 1$, and linear between its values $H(n) = \delta(n-1)$. The translates $H(t-k)$ generate all piecewise linear functions on unit intervals. Any function $f(t)$ in V_0 can be expressed as $\sum f(n-1)H(t-n)$. However $H(t)$ is *not orthogonal* to the neighboring hat $H(t-1)$. The product $H(t)H(t-1)$ is positive on the one interval $1 < t < 2$ where the hats overlap. Its integral (the inner product of the hats) is not zero.

We must work harder to find an orthogonal basis, and the eventual $\phi(t)$ will not have compact support. Or else we keep this non-orthogonal basis.

3. Discontinuous piecewise linear functions. Now $f(t)$ in V_0 may have a jump at each meshpoint $t = n$. There is a value $f(n_-)$ from the left and a value $f(n_+)$ from the right. The hat function is still in the space, but so is the box function! The spaces V_j are clearly shift-invariant and scale-invariant. If $f(t)$ is linear between integers (where it jumps), then $f(2t)$ is linear between half-integers (where it jumps).

There are two degrees of freedom at each meshpoint, the values $f(n_-)$ and $f(n_+)$. Therefore *two scaling functions* $\phi_1(t)$ and $\phi_2(t)$ are required for a shift-invariant basis. They can both be supported on the unit interval, and they can be orthogonal:

$$\phi_1(t) = \textbf{\textit{box function}} \quad \text{and} \quad \phi_2(t) = \textbf{\textit{sloping line}} = 1 - 2t.$$

The union of $\{\phi_1(t-k)\}$ and $\{\phi_2(t-k)\}$ is an orthonormal basis—which illustrates the idea behind "*multiwavelets*". The usual dilation equation for $\phi(t)$ becomes a vector equation for $\phi_1(t)$ and $\phi_2(t)$. The coefficients $c(k)$ in that equation are 2×2 matrices. The associated filter bank in Section 7.5 contains "*multifilters*".

4. Cubic splines. V_0 consists of piecewise cubic polynomials on unit intervals, with $f(t)$ and $f'(t)$ and $f''(t)$ continuous. The third derivative $f'''(t)$ may jump at the integers $t = n$, so the cubics are different in neighboring intervals. We have shift-invariance and scale-invariance, when V_1 contains the cubic splines on half-intervals. *This is the main point*: Approximating subspaces on regular meshes automatically fit the requirements for multiresolution.

The shortest cubic spline is a *B-spline*. It consists of different third-degree polynomials on the four unit intervals within $0 \le t \le 4$. The letter B stands for basis, but not for orthogonal basis. In complete analogy with the hat function, which is a linear spline, the scaling function $\phi(t)$ for the cubic splines cannot have compact support if we insist on orthogonality. An orthogonal basis $\{\phi(t-k)\}$ does exist, but it requires Fourier analysis to find it.

The cubic B-spline satisfies a dilation equation with very simple coefficients, proportional to $1, 4, 6, 4, 1$. But those coefficients do not lead to orthogonal filters. We can stay with these coefficients and go to biorthogonal filters (the best plan). Or we can orthogonalize, losing compact support and reaching a filter with infinitely many coefficients. Section 7.4 develops the theory of splines.

5. Daubechies functions. The search for orthogonal filter banks leads to the four coefficients of a "maxflat" lowpass filter. The response $C(\omega)$ has a double zero at the highest frequency $\omega = \pi$. This is maximal flatness, with four coefficients and orthogonality:

$$c(0),\ c(1),\ c(2),\ c(3) = 1+\sqrt{3},\ 3+\sqrt{3},\ 3-\sqrt{3},\ 1-\sqrt{3} \text{ times } 1/4\sqrt{2}.$$

Their sum is $\sqrt{2}$. Their sum of squares is unity. They are orthogonal to their double shifts, because $c(0)c(2)+c(1)c(3)=0$. From these coefficients Daubechies constructed $\phi(t)$ by solving the dilation equation

$$\phi(t)=\sqrt{2}\sum_{k=0}^{3} c(k)\,\phi(2t-k).$$

The solution comes in the next section. The zeroth space V_0 contains every $\phi(t-k)$. Those functions are an orthonormal basis. The rescaled functions $\phi(2^j t-k)$ span V_j.

This is our best description of the Daubechies spaces V_j, to give the dilation equation for $\phi(t)$. In Examples 1–4, we started with the spaces. In Example 5, Daubechies started with the coefficients and found $\phi(t)$—which produces the spaces. Either way, we have the scale-invariance and shift-invariance of multiresolution analysis.

The Dilation Equation

The space V_0 is contained in V_1. Therefore $\phi(t)$ is also in V_1. It must be a combination of the basis functions $2^{1/2}\phi(2t-k)$ for that subspace. The coefficients in the combination will be called $c(k)$. Bring the factor $2^{1/2}=\sqrt{2}$ outside:

$$\boxed{V_0 \subset V_1 \text{ means } \phi(t)=\sqrt{2}\sum_k c(k)\,\phi(2t-k).} \tag{6.11}$$

This is the dilation equation. It is a two-scale equation, involving t and $2t$. It is also called a *refinement equation*, because it displays $\phi(t)$ in the refined space V_1. That space has the finer scale $\Delta t=1/2$, and it contains $\phi(t)$ which has scale $\Delta t=1$.

To emphasize: The dilation equation is a direct consequence of $V_0 \subset V_1$. It is not an extra requirement! There will be a finite set of coefficients $c(0),\ldots,c(N)$ when $\phi(t)$ is supported on $[0,N]$. In general, $\phi(t)$ has infinite support and we need infinitely many $c(n)$.

To find $c(n)$, multiply the dilation equation (6.11) by $\sqrt{2}\phi(2t-n)$. Integrate and use orthogonality:

$$\sqrt{2}\int_{-\infty}^{\infty}\phi(t)\phi(2t-n)\,dt=c(n). \tag{6.12}$$

If $\phi(t)$ is the unit box and $\phi(2t)$ is the half-box, this gives $c(0)=\sqrt{2}/2$ and $c(1)=\sqrt{2}/2$. The dilation equation for the box function then has coefficients 1 and 1:

$$\phi(t)=\phi(2t)+\phi(2t-1). \tag{6.13}$$

From orthogonality of the basis $\{\phi(t-k)\}$ we have double-shift orthogonality of the dilation coefficients $c(k)$. And unit energy in $\phi(t)$ gives a unit vector of c's:

$$\text{Double-shift: } \sum c(k)c(k-2m)=\delta(m). \quad \text{Unit vector: } \sum |c(k)|^2=1 \tag{6.14}$$

For proof, multiply the dilation equations for $\phi(t)$ and $\phi(t-m)$ and integrate. Orthonormality of the ϕ's yields double-shift orthogonality of the c's:

$$\int_{-\infty}^{\infty}\phi(t)\phi(t-m)\,dt=\sum_k c(k)c(k-2m)=\delta(m). \tag{6.15}$$

The coefficients $c(k)$ go into an orthonormal filter bank! Starting with the spaces V_j in a multiresolution, the dilation equation has brought us back to filters—where the key matrix is $L = (\downarrow 2)C$. Double-shift orthogonality becomes $LL^T = I$. The rows of L contain the double shifts $L_{ij} = c(2i - j)$.

Box example:

$$L = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & & & & & \\ & & 1 & 1 & & & \\ & & & & 1 & 1 & \\ & & & & & & \cdots \end{bmatrix}$$

Daubechies example:

$$L = \begin{bmatrix} c(3) & c(2) & c(1) & c(0) & & & \\ & & c(3) & c(2) & c(1) & c(0) & \\ & & & & c(3) & c(2) & \cdots \\ & & & & & & \cdots \end{bmatrix}.$$

To end this section, we have to identify the wavelet spaces W_j.

The Wavelet Equation

The scaling functions $\phi(2^j t - k)$ are orthogonal at each scale separately. But $\phi(t)$ is not orthogonal to $\phi(2t)$. They are *not* orthogonal across scales; the level j must be fixed. The function $\phi(t)$ in V_0 is also in V_1 (the dilation equation). Orthogonality *across scales* comes from the wavelet subspaces W_j and their basis functions $w_{jk}(t)$. We study those now, from three starting-points:

1. The spaces W_j. **2.** The wavelets $w(t)$. **3.** The coefficients $d(k)$.

Use Method **1** if you have the V_j. Their differences yield the spaces W_j. Use Method **2** if you can identify the wavelets. Just shift and rescale. Use Method **3** if you have the numbers $c(k)$. The *alternating flip* yields $d(k) = (-1)^k c(N - k)$. Then $w(t)$ comes from the wavelet equation below, and W_j contains the combinations of $w(2^j t - k)$.

The box function gives an example in which all three approaches will work. We construct W_0 and $w(t)$ and the d's:

1. *From the subspaces*: V_0 contains constant functions on unit intervals, and V_1 contains constant functions on *half*-intervals. The space W_0 is in V_1 (therefore constant on half-intervals). It is orthogonal to V_0, so the integral over each full interval is zero. This fact produces the complementary subspace W_0, orthogonal to V_0 inside V_1:

$$W_0 = \{ \text{constants on half-intervals with } f(n) + f(n + 1/2) = 0 \}.$$

$V_0 \bigoplus W_0$ does give V_1. Combining *equal* values at n and $n + 1/2$ from V_0 with *opposite* values from W_0 gives *any* values $f(n)$ and $f(n + 1/2)$ for V_1.

2. *From the wavelets:* The important function in W_0 is the up-down square wave:

$$\textit{Haar wavelet} \quad w(t) = \begin{cases} 1 & \text{for } 0 \le t < \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

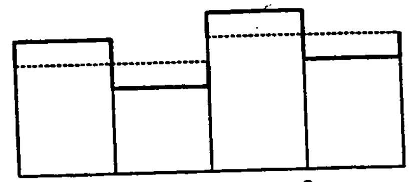


Figure 6.3: Two bases for V_1: Halfsize boxes $\phi(2t-k)$ or full boxes $\phi(t-k)$ plus up-down Haar wavelets $w(t-k)$.

This is orthogonal to the box function $\phi(t)$. It is orthogonal to translates of ϕ and also to its own translates (there is no overlap of $w(t)$ with $w(t-1)$). *More than that, multiresolution says that the wavelet $w(t)$ is orthogonal to rescalings of itself and to translates of rescalings:*

$$\int_{-\infty}^{\infty} w(t)w(2^j t-k)\,dt = 0 \text{ unless } j=k=0.$$

The translates of $w(t)$ span W_0. The translates of $w(2^j t)$ span W_j. Those wavelet spaces are orthogonal because $W_0 \subset V_j$ and $V_j \perp W_j$. (Exchange j and 0 if j is negative.) From orthogonal spaces we have orthogonal basis functions. Then completeness makes the whole orthonormal system $\{2^{j/2}w(2^j t-k)\}$ a basis for L^2.

3. *From the coefficients* $c(0)=c(1)=1/\sqrt{2}$: The flip construction gives $d(0)=1/\sqrt{2}$ and $d(1)=-1/\sqrt{2}$. Those coefficients go into the wavelet equation:

$$\boxed{\textit{Wavelet equation} \quad w(t)=\sqrt{2}\sum d(k)\,\phi(2t-k).} \tag{6.16}$$

This equation produces the wavelet directly from the scaling functions—*no equation to solve*! The wavelet is $w(t)=\phi(2t)-\phi(2t-1)$. This is a half-box minus a shifted half-box. It is the up-down square wave, which is Haar's wavelet.

Our final example, from Daubechies, starts with the four c's. Then the flip construction gives the four d's (to normalize, divide again by $4\sqrt{2}$):

$$d(0), d(1), d(2), d(3) = 1-\sqrt{3}, -(3-\sqrt{3}), 3+\sqrt{3}, -(1+\sqrt{3}).$$

Their sum is zero. Their sum of squares (normalized) is 1. They are orthogonal to their double shifts, because the c's are. The wavelet equation gives the Daubechies wavelet $w(t)$, which has no simple formula. *The orthogonality to $w(t-k)$ and $\phi(t-k)$ is only known indirectly*—from the double-shift orthogonality of the d's. The structure of multiresolution gives crucial information that we cannot find in a table of integrals.

The actual construction of $\phi(t)$ and $w(t)$, and the drawing of their graphs, is immediately ahead.

Example 6.2. (Strange but beautiful.) Suppose $\phi(t)$ is the delta function $\delta(t)$. This is not in L^2 but continue anyway. The space V_0 contains combinations $\sum a(n)\delta(t-n)$ of *delta functions at the integers*. What orthogonal wavelet $w(t)$ goes with this scaling function $\delta(t)$?

By scale invariance, V_1 contains $\delta(2t-n)$. The spikes for V_1 are at $t=0,\pm\frac{1}{2},\pm 1,\ldots$. Since V_0 holds the delta functions at integers, W_0 contains delta functions *at the midpoints* $t=n+\frac{1}{2}$.

he integers and the midpoints combine to give $V_0 \oplus W_0 = V_1$. *The wavelet is the delta function* $t\ t = \frac{1}{2}$.

Similarly, V_j contains delta functions at $t = n/2^j$. W_j contains delta functions at the mid-oints $(n + \frac{1}{2})/2^j$. What is W_{-1}? Its delta functions are at $t = (n + \frac{1}{2})/2^{-1} = 2n + 1$. These re odd integers ± 1, ± 3, ± 5, ... The spacing between them is 2 as expected. Then W_{-2} has elta functions at ± 2, ± 6, ± 10, ... with spacing 4. The union of all W_j has delta functions at ll binary points.

The dilation equation for the delta function is $\delta(t) = 2\delta(2t)$. The only nonzero coefficient ; $h(0) = 1$. *The filter is the identity.* The wavelet equation with only one term is $w(t) = \delta(2t - 1) = \delta(t - \frac{1}{2})$. This confirms what we found, that W_0 contains delta functions at all iidpoints between integers. Notice! An odd number of coefficients (one) means that $N = 0$:ven). *The alternating flip must shift by an odd integer, for double-shift orthogonality.* So the onzero highpass coefficient was $d(1)$ not $d(0)$.

To linger one last second on this trivial great example, the double-shift matrices from the low nd high channels are $L = (\downarrow 2)$ and $B = (\downarrow 2)(\text{ delay })$.

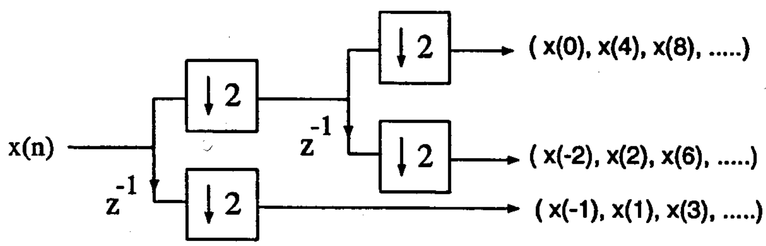


Figure 6.4: The lazy filter $H = I$ leads to delta functions.

he Scaling Function is Supported on $[0, N]$

remarkable feature of $\phi(t)$ is that it is zero outside the interval $0 \le t \le N$. This could never appen to a ***one-scale*** difference or differential equation (homogeneous). The solutions would e combinations of λ^n and $e^{\lambda t}$, and only occasionally zero. The compact support of $\phi(t)$ comes rom the two scales in the dilation equation

$$\phi(t) = \sum_{k=0}^{N} 2h(k)\,\phi(2t - k). \tag{6.17}$$

heorem 6.1 *The scaling function $\phi(t)$ is supported on the interval* $[0, N]$.

roof. Suppose we know that the support is a finite interval $[a, b]$. Then $\phi(2t)$ is supported on $\frac{a}{2}, \frac{b}{2}]$. The shifted function $\phi(2t - k)$ is supported on $[\frac{a+k}{2}, \frac{b+k}{2}]$. The index k goes from zero N, so the right side of the dilation equation is supported between $\frac{a}{2}$ and $\frac{b+N}{2}$. Comparing with he left side,

$$[a, b] = \left[\frac{a}{2}, \frac{b+N}{2}\right] \quad \text{leads to} \quad a = 0 \text{ and } b = N.$$

This section will do part of each job, the ***recursive part***. This shows how multiresolution (for functions) connects to subband filtering (for vectors). The three parts that we can do immediately are:

1. Compute a_{jk} and b_{jk} recursively from $a_{j+1,k}$ (and vice versa).

2. Set up a recursion (the cascade algorithm) to construct $\phi(t)$.

3. Prove orthogonality for $\phi_{jk}(t)$ and $w_{jk}(t)$ from orthogonality of $\boldsymbol{c}$'s and $\boldsymbol{d}$'s.

Those are the three subsections. Later we have to initialize the recursion in 1, execute and study the cascade algorithm in 2, and derive other properties in 3.

Wavelet Coefficients by Recursion

Suppose $f_1(t)$ is in V_1. It is a combination of the basis functions $\sqrt{2}\,\phi(2t-k)$. These functions $\phi_{1k}(t)$ are at level 1. Multiresolution splits this level into $V_1 = V_0 \oplus W_0$, so $f_1(t)$ is also a combination of the basis functions for V_0 and W_0. Those basis functions are $\phi_{0k}(t) = \phi(t-k)$ and $w_{0k}(t) = w(t-k)$:

$$\begin{aligned} \sum a_{1k}\,\phi_{1k}(t) &= \sum a_{0k}\,\phi_{0k}(t) + \sum b_{0k}\,w_{0k}(t) \\ &= \sum a_{0k}\,\phi(t-k) + \sum b_{0k}\,w(t-k). \end{aligned} \tag{6.19}$$

We are computing a change of basis. Given the coefficients $a_{1k}(t)$ in the V_1 basis, we want the coefficients a_{0k} and b_{0k} in the $V_0 \oplus W_0$ basis. The same step will apply at every level. It takes us from the coefficients $a_{j+1,k}$ in the basis for V_{j+1}, to the coefficients a_{jk} and b_{jk} in the bases for V_j and W_j. This is the recursion that makes the wavelet transform fast.

We will suppose that these bases are ***orthonormal***. Later in this section we prove this property (assuming the cascade algorithm uses orthogonal filters and converges). Orthonormality makes the formulas easy and it makes the inverse easy. Section 6.5 will derive the biorthogonal recursion, when orthogonality is not assumed.

To find the recursion, shift equation (6.18) by k and set $n = \ell - 2k$:

$$\begin{aligned} &\textbf{\textit{Dilation equation: }} \phi(t-k) = \sum \sqrt{2}\,c(n)\,\phi(2t-2k-n) = \sum c(\ell-2k)\,\phi_{1\ell}(t) \\ &\textbf{\textit{Wavelet equation: }} w(t-k) = \sum \sqrt{2}\,d(n)\,\phi(2t-2k-n) = \sum d(\ell-2k)\,\phi_{1\ell}(t) \end{aligned} \tag{6.20}$$

Multiply by $f_1(t)$ and integrate with respect to t. Since the basis functions are orthonormal, the integral gives the coefficients of $f_1(t)$ in each basis:

$$a_{0k} = \sum c(\ell-2k)\,a_{1\ell} \quad \text{and} \quad b_{0k} = \sum d(\ell-2k)\,a_{1\ell}. \tag{6.21}$$

This is the key recursion. It is the action of a filter bank, which inputs $a_{1\ell}$ and outputs a_{0k} and b_{0k}. But we have to watch indices, because an ordinary convolution would be $\sum c(k-\ell)\,a_{1\ell}$ and downsampling would give $\sum c(2k-\ell)\,a_{1\ell}$. There is a time reversal between this filter $\boldsymbol{C}$ and the *transpose* filter $\boldsymbol{C}^T$ that appears in the recursion (6.21):

$$c^T(n) = c(-n) \quad \text{and} \quad d^T(n) = d(-n). \tag{6.22}$$

How do we know that the support is a finite interval in the first place? From the cascade algorithm. The box function $\phi^{(0)}(t)$ is supported on $[0, 1]$. When this box is substituted into the right side of the dilation equation, the function $\phi^{(1)}(t)$ that comes out has support $[0, \frac{1+N}{2}]$. Then $\phi^{(1)}(t)$ is substituted into the right side and the result $\phi^{(2)}(t)$ is zero outside $[0, \frac{1+3N}{4}]$. The limiting function $\phi(t)$ is certain to be zero outside $[0, N]$. This cascade is studied in the next section.

It will be useful to reach the same conclusion based on the Fourier transform (Section 6.4). That argument can assume less about the filter coefficients. We mention that there are never gaps where $\phi(t)$ is zero on an interval inside $[0, N]$. And if the highpass coefficients $h_1(k)$ run from $k = 0$ to $k = \widetilde{N}$, then the wavelet $w(t) = \sum 2h_1(k)\phi(2t - k)$ has support $[0, \frac{1}{2}(N + \widetilde{N})]$. The last term $\phi(2t - \widetilde{N})$ is zero after $2t - \widetilde{N}$ reaches N.

Problem Set 6.1

1. Explain why the scaling requirement, that $f(t)$ is in V_j if and only if $f(2t)$ is in V_{j+1}, can be restated as $\widehat{f}(\omega)$ is in $\widehat{V}_j$ if and only if $\widehat{f}(2\omega)$ is in $\widehat{V}_{j-1}$. Here $\widehat{V}_j$ is the space of Fourier transforms of functions in V_j.
2. For the space V_0 of piecewise constant functions in Example 1, show that the only shift-invariant basis $\phi(t - k)$ contains box functions. What is the corresponding statement about allpass FIR filters?
3. For piecewise constants, show that $f(t)$ is in L^2 if and only if $f(n)$ is in l^2.
4. Find 2 by 2 matrices $c(0)$ and $c(1)$ so that the box function $\phi_1(t)$ and sloping line $\phi_2(t) = 1-2t$ in Example 3 satisfy

$$\begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix} = c(0)\begin{bmatrix} \phi_1(2t) \\ \phi_2(2t) \end{bmatrix} + c(1)\begin{bmatrix} \phi_1(2t-1) \\ \phi_2(2t-1) \end{bmatrix}.$$

5. If $f(t)$ is in V_0 and $g(t)$ is in V_1, why is it generally false that $g(t) - f(t)$ is in W_1?
6. What multiresolution requirements are violated if W_j consists of all multiples of $\cos(2^j t)$?

6.2 Wavelets from Filters

The previous section reached the dilation equation and the wavelet equation:

$$\phi(t) = \sum \sqrt{2}\, c(n)\, \phi(2t - n) \quad \text{and} \quad w(t) = \sum \sqrt{2}\, d(n)\, \phi(2t - n). \tag{6.18}$$

Those equations are the crucial connections between wavelets and filters. Historically, their development was separate. Now you have to see them together. The lowpass filter $c(0), \ldots, c(N)$ determines the scaling function $\phi(t)$. Then the highpass coefficients produce the wavelets.

Working with $\phi(t)$ and $w(t)$, we really have three basic jobs:

1. Compute the coefficients in $f_j(t) = \sum_k a_{jk}\phi_{jk}(t)$ and $f(t) = \sum_j \sum_k b_{jk} w_{jk}(t)$.
2. Construct $\phi(t)$ by actually solving the dilation equation.
3. Connect the properties of $\phi(t)$ and $w(t)$ to properties of the c's and d's.

Going between levels of a multiresolution is subband filtering with C^T and D^T:

Theorem 6.2 *A function $\sum a_{j+1,\ell}\,\phi_{j+1,\ell}(t)$ in the space $V_{j+1} = V_j \oplus W_j$ has coefficients a_{jk} and b_{jk} in the new orthonormal basis $\{\phi_{jk}(t), w_{jk}(t)\}$:*

$$\boxed{a_{jk} = \sum_{\ell} c(\ell - 2k)\, a_{j+1,\ell} \quad \text{and} \quad b_{jk} = \sum_{\ell} d(\ell - 2k)\, a_{j+1,\ell}.} \tag{6.23}$$

In vector notation this is $\boldsymbol{a}_j = (\downarrow 2)\, C^T\, \boldsymbol{a}_{j+1}$ and $\boldsymbol{b}_j = (\downarrow 2)\, D^T\, \boldsymbol{a}_{j+1}$. The pyramid is

$$\begin{tikzcd}
a_{j+1} \arrow[r, "C^T"] \arrow[dr, "D^T"'] & a_j \arrow[r, "C^T"] \arrow[dr, "D^T"'] & a_{j-1} \arrow[r] \arrow[dr] & \cdots & a_1 \arrow[r, "C^T"] \arrow[dr, "D^T"'] & a_0 \\
& b_j & b_{j-1} & & & b_0
\end{tikzcd}$$

Proof. For $j = 0$, formula (6.23) is (6.21). The extension to every j comes from the dilation equation. Again $n = \ell - 2k$:

$$2^{j/2}\,\phi(2^j t - k) = 2^{j/2} \sum \sqrt{2}\, c(n)\, \phi(2^{j+1}t - 2k - n) = \sum c(\ell - 2k)\, \phi_{j+1,\ell}(t). \tag{6.24}$$

The wavelet equation has $\boldsymbol{d}$ in place of $\boldsymbol{c}$. The inner products of these equations with $f(t)$ give the recursions (6.23) for the coefficients a_{jk} and b_{jk}.

Now go in the opposite direction. Change from the basis $\{\phi_{jk}(t), w_{jk}(t)\}$ back to the basis $\{\phi_{j+1,\ell}(t)\}$. Since the bases are orthonormal, the inverse operation is given by the transpose.

Theorem 6.3 *$a_{j+1,\ell}$ comes from a_{jk} and b_{jk} by a synthesis filter bank:*

$$\boxed{a_{j+1,\ell} = \sum c(2k - \ell)\, a_{jk} + d(2k - \ell)\, b_{jk}.} \tag{6.25}$$

The inverse pyramid is the fast inverse wavelet transform:

$$\begin{tikzcd}
a_0 \arrow[r, "C"] & a_1 \arrow[r, "C"] & a_2 & \cdots & a_j \arrow[r, "C"] & a_{j+1} \\
b_0 \arrow[ur, "D"'] & b_1 \arrow[ur, "D"'] & & & b_j \arrow[ur, "D"'] &
\end{tikzcd}$$

Lowpass Iteration and the Cascade Algorithm

We begin the solution of the dilation equation. Our goal is to construct the scaling function $\phi(t)$. The only inputs are the filter coefficients $c(0), \ldots, c(N)$. The first solution method we propose is the ***cascade algorithm***.

Start the cascade with $\phi^{(0)}(t) =$ box function on $[0, 1]$. *Iterate the lowpass filter:*

$$\phi^{(i+1)}(t) = \sum_n \sqrt{2}\, c(n)\, \phi^{(i)}(2t - n) = \sum_n 2\, h(n)\, \phi^{(i)}(2t - n). \tag{6.26}$$

The algorithm works with functions in continuous time. Those functions are piecewise constant and the pieces become shorter (their length is 2^{-i}). If $\phi^{(i)}(t)$ converges suitably to a limit $\phi(t)$, then this limit function solves the dilation equation.

Notice the two time scales, t and $2t$, which come from the continuous form of downsampling. In place of $(\downarrow 2)\,\phi(n) = \phi(2n)$, we have $(\downarrow 2)\,\phi(t) = \phi(2t)$. The cascade algorithm is really iteration with the filter matrix $M = (\downarrow 2)\,2H$—as we will see in detail. It is an infinite iteration, and our final formula for $\phi(t)$ will involve an infinite product.

It is easy to associate a continuous-time function $x(t)$ with a discrete-time vector $x(n)$. The function takes the value $x(n)$ over the n^{th} time interval. That is the interval $n \le t < n+1$. Thus the constant vector $x = (\ldots, 1, 1, 1, \ldots)$ produces the constant function $x(t) \equiv 1$. The impulse $x = (\ldots, 0, 1, 0, \ldots)$ produces the standard *box function*. In general $x(t)$ is *piecewise constant:* $x(t) = x(n)$ on the interval $n \le t < n+1$.

The iterations start from the box function $\phi^{(0)}(t)$. There are two steps in each iteration—*filtering* and *rescaling*. Suppose the filter coefficients are $h(0) = 2/3$ and $h(1) = 1/3$. Filtering the input gives $\frac{2}{3}\phi^{(0)}(t) + \frac{1}{3}\phi^{(0)}(t-1)$. Then rescaling t to $2t$ compresses the graph. *To maintain a constant area we multiply the height by* 2:

$$\phi^{(1)}(t) = \tfrac{4}{3}\phi^{(0)}(2t) + \tfrac{2}{3}\phi^{(0)}(2t-1).$$

Filtering and rescaling one box produces two half-width boxes of height $\frac{4}{3}$ and $\frac{2}{3}$. That iteration step preserves the area (=1). Now filter and rescale $\phi^{(1)}(t)$. The two half-boxes become four quarter-boxes, from $\phi^{(2)}(t) = \frac{4}{3}\phi^{(1)}(2t) + \frac{2}{3}\phi^{(1)}(2t-1)$. The first quarter-box has height $\frac{16}{9}$. That height is multiplied by $\frac{4}{3}$ at every iteration!

We wish we could say that the iterations $\phi^{(i)}(t)$ are converging. Their limit $\phi(t)$ would satisfy the dilation equation $\phi(t) = \frac{4}{3}\phi(2t) + \frac{2}{3}\phi(2t-1)$. In some weak sense, this may be true. In a pointwise sense at $t = 0$, the functions $\phi^{(i)}(0)$ diverge because of $(4/3)^i$. The coefficients $2/3$ and $1/3$ illustrate the iteration process, but not its convergence.

We want to see that process also by algebra. It is clearest if we ignore the rescaling and just execute the filtering with coefficients $h(k)$. The heights of the boxes would be $\frac{2}{3}, \frac{1}{3}$, and then $\frac{4}{9}, \frac{2}{9}, \frac{2}{9}, \frac{1}{9}$. In the z-domain, this corresponds to

$$H(z) = \tfrac{2}{3} + \tfrac{1}{3}z^{-1} \quad \text{and} \quad H(z^2)H(z) = \tfrac{4}{9} + \tfrac{2}{9}z^{-1} + \tfrac{2}{9}z^{-2} + \tfrac{1}{9}z^{-3}. \tag{6.27}$$

The actual time intervals go from length 1 to $\frac{1}{2}$ to $\frac{1}{4}$. The actual graph heights are doubled at each step, to preserve area. But the essential point is the product $H(z^2)H(z)$. After three steps, the iteration will produce $H^{(3)}(z) = H(z^4)H(z^2)H(z)$. After i steps we have

$$H^{(i)}(z) = \prod_{k=0}^{i-1} H(z^{2^k}). \tag{6.28}$$

This product is the z-domain equivalent of iterating the lowpass filter $H(z)$. The values of $\phi^{(i)}(t)$—the heights of the graph after i iterations—are the coefficients of $2^i H^{(i)}(z)$. That factor 2^i accounts for the height-doublings that preserve area, when the time intervals for $\phi^{(i)}(t)$ become 2^{-i}.

You may ask, why not choose the usual averaging filter as a first example? Let me show you why. The averaging coefficients are $h(0) = h(1) = \frac{1}{2}$. The first step of the iteration, with coefficients $2h(0) = 2h(1) = 1$, is

$$\phi^{(1)}(t) = \phi^{(0)}(2t) + \phi^{(0)}(2t-1).$$

From the box function $\phi^{(0)}(t)$ this produces the same box: $\phi^{(1)}(t) = \phi^{(0)}(t)$.

The output equals the input. The iteration process converges immediately. We have found the scaling function! In general $\phi(t)$ is the limit of the sequence $\phi^{(i)}(t)$, when that limit exists as $i \to \infty$. Here $\phi^{(0)} = \phi^{(1)}$ and the box function is a "fixed point" of the iteration. When we filter and rescale $\phi(t)$ we get back $\phi(t)$, because the sum of two half-length boxes is the original box:

$$\boxed{\textit{Box:}\quad \phi(t) = \phi(2t) + \phi(2t-1).} \tag{6.29}$$

The z-domain equivalent is a product built from

$$H(z) = \tfrac{1}{2} + \tfrac{1}{2}z^{-1}.$$

Please notice that we *do not square* this function. $H^{(2)}(z)$ is $H(z^2)H(z)$:

$$H^{(2)}(z) = (\tfrac{1}{2} + \tfrac{1}{2}z^{-2})(\tfrac{1}{2} + \tfrac{1}{2}z^{-1}) = \tfrac{1}{4} + \tfrac{1}{4}z^{-1} + \tfrac{1}{4}z^{-2} + \tfrac{1}{4}z^{-3}. \tag{6.30}$$

After i iterations, $H^{(i)}(z)$ will have 2^i coefficients all equal to 2^{-i}. After rescaling, this still corresponds to the box function.

Now use three filter coefficients $h = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. The box $\phi^{(0)}(t)$ produces *three* half-boxes in

$$\phi^{(1)}(t) = \tfrac{1}{2}\phi^{(0)}(2t) + \phi^{(0)}(2t-1) + \tfrac{1}{2}\phi^{(0)}(2t-2).$$

Then there are *seven* quarter-boxes in $\phi^{(2)}(t)$. Rescaling prevents the support interval from becoming long. The limiting interval is $0 \le t < 2$.

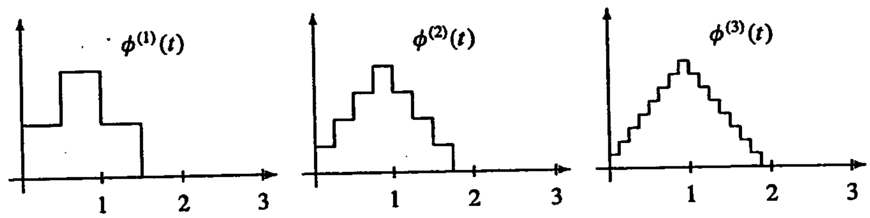


Figure 6.5: The cascade algorithm for $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ converges to the hat function.

A reasonable guess for the limiting function $\phi(t)$ is the *hat function*. This is piecewise linear, going up to $\phi(1) = 1$ and down to $\phi(2) = 0$. We verify that the hat function is a fixed point of the iteration. Filtering and rescaling leaves this scaling function $\phi(t)$ unchanged:

$$\phi(t) = \tfrac{1}{2}\phi(2t) + \phi(2t-1) + \tfrac{1}{2}\phi(2t-2). \tag{6.31}$$

Notice how the coefficients $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ are doubled. The hat function is a combination of three narrower hats. For future reference, we note the different properties of these examples:

1. $H(z) = \frac{2}{3} + \frac{1}{3}z^{-1}$ is not zero at $z = -1$, corresponding to $\omega = \pi$. The iterations fail to converge.

2. $H(z) = \frac{1}{2} + \frac{1}{2}z^{-1}$ is zero at $z = -1$. The iterations converge. The filter $H(z)H(z^{-1})$ is halfband: no even powers except the constant term. The box function is orthogonal to its translates.

3. $H(z) = \frac{1}{4} + \frac{1}{2}z^{-1} + \frac{1}{4}z^{-2}$ is zero (twice) at $z = -1$. The iterations converge. The filter $H(z)H(z^{-1})$ is *not* halfband. It contains the even powers z^2 and z^{-2}. The hat function $\phi(t)$ is *not* orthogonal to $\phi(t-1)$.

We must quickly emphasize that a zero at $z = -1$ (which is $\omega = \pi$ in the frequency domain) does not guarantee the convergence of $\phi^{(i)}(t)$. But without that zero in the filter response, strong convergence has no chance.

Similarly, a halfband filter does not guarantee that $\phi(t)$ is orthogonal to its translates. But without that halfband property of $H(z)H(z^{-1})$, orthogonality has no chance. Section 7.2 will further indicate those connections; they are not quite two-way implications:

Convergence of $\phi^{(i)}(t)$ to $\phi(t)$ needs $H = 0$ at $z = -1$.
Orthogonality of $\phi(t-k)$ needs $H(z)H(z^{-1})$ to be halfband

Orthogonal Functions from Orthogonal Filters

When the filter bank is orthonormal in discrete time, we hope for orthogonal basis functions in continuous time. All wavelets $w(2^j t)$ should be orthogonal to the scaling functions $\phi(t-k)$. Furthermore, the wavelets $w(2^j t - k)$ should be mutually orthogonal and the scaling functions $\phi(t-k)$ should be mutually orthogonal. Note that $\phi(t)$ is *not* orthogonal to $\phi(2t)$.

Theorem 6.4 *Assume that the cascade algorithm converges: $\phi^{(i)}(t) \to \phi(t)$ uniformly in t. If the coefficients $c(k)$ and $d(k)$ come from an orthonormal filter bank, so they have double-shift orthogonality, then*

1. *The scaling functions $\phi(t-n)$ are orthonormal to each other:*

$$\int_{-\infty}^{\infty} \phi(t-n)\phi(t-m)\,dt = \delta(m-n).$$

2. *The scaling functions are orthogonal to the wavelets:*

$$\int_{-\infty}^{\infty} \phi(t-m)w(t-n)\,dt = 0.$$

3. *The wavelets $w_{jk}(t) = 2^{j/2}w(2^j t - k)$ at all scales are orthonormal:*

$$\int_{-\infty}^{\infty} w_{jk}(t)w_{JK}(t)\,dt = \delta(j-J)\delta(k-K).$$

Proof of 1: The box functions $\phi^{(0)}(t-k)$ are certainly orthonormal (because nonoverlapping). We will show that when $\phi^{(i)}(t-k)$ are orthogonal, the next iterates $\phi^{(i+1)}(t-k)$ are also orthonormal. Then the limits $\phi(t-k)$ are orthonormal.

The induction step from i to $i+1$ assumes that the $\phi^{(i)}(t-k)$ are orthonormal, and sets $l = m - n$:

$$\begin{aligned}
&\int \phi^{(i+1)}(t-m)\phi^{(i+1)}(t-n)\,dt \\
&= 2\int \Big(\sum c(k)\phi^{(i)}(2t-2m-k)\Big)\Big(\sum c(k)\phi^{(i)}(2t-2n-k)\Big)\,dt \qquad (6.32)\\
&= \int \Big(\sum c(k)\phi^{(i)}(2t-2m-k)\Big)\Big(\sum c(k-2l)\phi^{(i)}(2t-2m-k)\Big)\,2dt \\
&= \sum c(k)c(k-2l) = \delta(l) = \delta(m-n).
\end{aligned}$$

The crucial step came in the last line, when we used the orthogonality of the row $[c(0)\cdots c(N)]$ to its double shifts. These are rows of $L = (\downarrow 2)C$. The orthogonality is in the statement $LL^T = I$. Equivalently, it is in the statement that $|\sum c(k)e^{-ik\omega}|^2$ is a normalized halfband filter: no even powers except the constant term 1.

Note the important point! Orthogonality of wavelets came from orthogonality of filters. When the infinite iterations converge, the limits retain orthogonality. This holds at each scale level j. In $\int \phi(2^jt-m)\phi(2^jt-n)\,dt$, we replace 2^jt by T. Orthogonality *does not hold* between scaling functions at different levels. Certainly, $\phi(t)$ is not orthogonal to all $\phi(2t-n)$, or the dilation equation would require $\phi \equiv 0$.

Proof of 2: Repeat the integration steps above for ϕ times w:

$$\begin{aligned}
&\int \phi(t-m)w(t-n)\,dt \\
&= \int \Big(\sum c(k)\,\sqrt{2}\,\phi(2t-2m-k)\Big)\Big(\sum d(k)\,\sqrt{2}\,\phi(2t-2n-k)\Big)\,dt \\
&= \cdots = \sum c(k)d(k-2l) = 0.
\end{aligned}$$

Always, $l = m - n$. The last step uses the orthogonality of the rows of $L = (\downarrow 2)C$ to the rows of $B = (\downarrow 2)D$. Again the double shift is essential. It is false that *all* rows of C and D are orthogonal.

The matrix form of this double-shift orthogonality is $LB^T = \mathbf{0}$. It comes from the alternating flip. That choice always produces double-shift orthogonality of d's to c's, but it does not by itself make $w(t)$ orthogonal to $\phi(t)$. To reach the end of part **2**, *we needed part* **1**—orthogonality between the ϕ's.

Proof of 3: The orthogonality of wavelets $w_{jk}(t)$ at the same scale level (the same j) is proved as in parts **1** and **2**:

$$\int_{-\infty}^{\infty} w(t-m)w(t-n)\,dt = \cdots = \sum d(k)d(k-2l) = \delta(l) = \delta(m-n).$$

Again continuous time orthogonality follows from discrete time orthogonality. This is not $DD^T = I$. It is $BB^T = I$, with double shifts in the rows of $B = (\downarrow 2)D$.

The orthogonality of wavelets at different scale levels (different j) is immediate from the rules of multiresolution. Suppose $j < J$. Then W_J is orthogonal to V_J by part **2**. But W_j is contained in V_{j+1} and therefore in V_J. So W_j is orthogonal to W_J. This proves the orthogonality theorem.

Final note: It was convenient to start from the box function $\phi^{(0)}(t)$, which is orthogonal to its translates. Then an orthogonal filter bank maintains this orthogonality to translates. Other starting functions will lead to the same fixed point $\phi(t)$, or at least to a multiple $c\phi(t)$—*if strong convergence holds.*

In general, convergence can be "*weak*" or "*strong*". For weak convergence, the functions $\phi^{(i)}(t)$ can oscillate faster and faster. You would not call this convergence. But the *integral* of $\phi^{(i)}(t)$ converges to the integral of $\phi(t)$, on every fixed interval $[0, T]$. (Integration controls the oscillations.) In the convergence that we assumed, $\phi^{(i)}(t)$ approaches $\phi(t)$ at every point.

There is a better starting function $\phi^{(0)}(t)$ than the box. The constant value $\phi^{(0)}(n)$ on each interval $n \leq t < n+1$ can be the *correct* $\phi(n)$. The values of ϕ are filled in at half-integers and quarter-integers by the iterations $\phi^{(1)}(t)$ and $\phi^{(2)}(t)$. The graph of $\phi(t)$ appears, 2^i points at a time. We stop when we have enough points for the printer to connect into a continuous graph.

The next section explains how to start with the correct values of $\phi(n)$ at the integers.

Problem Set 6.2

1. For the filter with $h(0) = h(1) = \frac{1}{2}$ and any $\phi^{(0)}(t)$, describe and draw $\phi^{(i)}(t)$.
2. If $H(z)$ is a polynomial of degree N, what is the degree of $H(z^2)H(z)$? What is the degree of $H^{(i)}(z) = \prod_{k=0}^{i-1} H(z^{2^k})$?

   Rescaling will replace z by $z^{1/2}$. After i steps, the degree is divided by 2^i. Show that the degree of $H^{(i)}(z^{1/2^i})$ approaches N as $i \to \infty$.
3. With coefficients $h(0), \ldots, h(N)$, the support interval of $\phi^{(i)}(t)$ grows to $[0, N]$. What happens if $\phi^{(0)}(t)$ is a box on $[0, 2N]$?
4. The unit area of the box is preserved if and only if $h(0) + \cdots + h(N) = 1$. Are negative coefficients allowed?
5. Suppose the filter coefficients $h(k)$ are $\frac{1}{2}, 0, 0, \frac{1}{2}$. Starting from the box function, take one step of the cascade algorithm and draw $\phi^{(1)}(t)$. Then take a second step and draw $\phi^{(2)}(t)$. Describe $\phi^{(i)}(t)$—on what fraction of the interval $[0, 3]$ does $\phi^{(i)}(t) = 1$?
6. Suppose the only filter coefficient is $h(0) = 1$. Starting from the box function $\phi^{(0)}(t)$, draw the graphs of $\phi^{(1)}(t)$ and $\phi^{(2)}(t)$. In what sense does $\phi^{(i)}(t)$ converge to the delta function $\delta(t)$? To verify the dilation equation $\delta(t) = 2\delta(2t)$, multiply by any smooth $f(t)$ and compare the integrals of both sides.
7. Suppose $\phi^{(0)}(t)$ is a stretched box of unit area: $\phi^{(0)}(t) = 1/2$ for $0 \leq t < 2$. Draw the graphs of $\phi^{(1)}(t)$ and $\phi^{(2)}(t)$ when $h(0) = h(1) = 1/2$. On what interval is $\phi^{(i)}(t)$ nonzero? What is the limit $\phi(t)$?
8. Suppose $\phi^{(0)}(t)$ is the Haar wavelet with *zero area*:

   $$\phi^{(0)}(t) = 1 \text{ for } 0 \leq t < 1/2 \text{ and } \phi^{(0)}(t) = -1 \text{ for } 1/2 \leq t < 1.$$

   With $h(0) = h(1) = 1/2$, draw the graphs of $\phi^{(1)}(t)$ and $\phi^{(2)}(t)$. The sequence $\phi^{(i)}(t)$ converges "weakly" to what multiple $c\phi(t)$?

6.3 Computing the Scaling Function by Recursion

The main point of this section can be stated in three sentences. Then you can follow through on the details, or look ahead for the matrices $m(0)$ and $m(1)$:

The dilation equation gives $\phi(0), \phi(1), \ldots$ as the eigenvector of a matrix $m(0)$.
Then $\phi(t)$ at t = half-integers comes from multiplying by a matrix $m(1)$.
Then $\phi(t)$ at every dyadic t comes by recursion. Each step uses $m(0)$ or $m(1)$.

The scaling function is created recursively. This section gives the rule.

The dilation equation is easiest with only *two coefficients* ($N = 1$). Then $m(0) = 2h(0)$ and $m(1) = 2h(1)$ are scalars not matrices. The two-coefficient dilation equation is

$$\boxed{\phi(t) = m(0)\,\phi(2t) + m(1)\,\phi(2t-1).} \tag{6.33}$$

The solution will be zero outside the interval $0 \le t < 1$. Inside that interval, set $t = 0$ to find $\phi(0)$ and $t = \frac{1}{2}$ to find $\phi(\frac{1}{2})$:

$$\phi(0) = m(0)\,\phi(0) \quad \text{and} \quad \phi\left(\tfrac{1}{2}\right) = m(1)\,\phi(0).$$

Now set $t = \frac{1}{4}$ and $\frac{3}{4}$, then $\frac{1}{8}$ and $\frac{5}{8}$, and onward through all the dyadic points $t = n/2^i$. Directly from the equation you find

$$\begin{aligned}
\phi\left(\tfrac{1}{4}\right) &= m(0)\phi\left(\tfrac{1}{2}\right) \quad \text{and} \quad \phi\left(\tfrac{3}{4}\right) = m(1)\phi\left(\tfrac{1}{2}\right)\\
\phi\left(\tfrac{1}{8}\right) &= m(0)\phi\left(\tfrac{1}{4}\right) \quad \text{and} \quad \phi\left(\tfrac{5}{8}\right) = m(1)\phi\left(\tfrac{1}{4}\right)\\
\phi\left(\tfrac{3}{8}\right) &= m(0)\phi\left(\tfrac{3}{4}\right) \quad \text{and} \quad \phi\left(\tfrac{7}{8}\right) = m(1)\phi\left(\tfrac{3}{4}\right).
\end{aligned}$$

Each new value comes from multiplying a previous value by $m(0)$ or $m(1)$. At each time t, the right side of equation (6.33) has only *one* nonzero term. Thus $\phi(\frac{3}{4})$ equals $m(1)\phi(\frac{1}{2})$ which is $m(1)m(1)\phi(0)$.

At the next step $\phi(\frac{3}{8})$ equals $m(0)m(1)m(1)\phi(0)$. The key is in the order of $m(0)$ and $m(1)$. *It is the same order as in the binary expansions* $\frac{3}{4} = 0.11$ *and* $\frac{3}{8} = 0.011$. At any point $t = n/2^i$, the solution $\phi(t)$ has i factors:

$$\boxed{\text{If} \quad t = 0.01101 \quad \text{in base 2, then} \quad \phi(t) = m(0)m(1)m(1)m(0)m(1)\phi(0).}$$

We have now solved the two-coefficient dilation equation at all dyadic points.

Admittedly, the restriction to two coefficients looks severe. The pattern is correct and important, but two numbers $m(0)$ and $m(1)$ are not enough. The only normal case is $m(0) = m(1) = 1$, when we get the box function. For $m(0) = \frac{4}{3}$ and $m(1) = \frac{2}{3}$, the first equation becomes $\phi(0) = \frac{4}{3}\phi(0)$. This produces a singularity of $\phi(t)$ at all dyadic points.

We will not pursue that example here, because there is a more valuable application — which reduces $N+1$ coefficients to two. This is the familiar step of reducing a high-order equation to a low-order system. For differential equations that produces a matrix, as in the system $u' = Au$. For dilation equations the reduction will produce *two matrices* $m(0)$ and $m(1)$. The dilation equation will become a ***two-coefficient matrix equation***. *The recursion will not change*, except it has vectors and matrices.

Vector Form of the Dilation Equation

The $N+1$ coefficients in the dilation equation are $\sqrt{2}c(k) = 2h(k)$:

$$\phi(t) \;=\; 2\sum_{k=0}^{N} h(k)\,\phi(2t-k). \tag{6.34}$$

Outside the interval $0 \le t < N$, we want and expect $\phi(t) \equiv 0$. Inside that interval, substitute $t = 0, t = 1, \ldots, t = N-1$ to determine $\phi(t)$ at the integers. You will see again the crucial

fact that even k goes with even $2t-k$, and odd k goes with odd $2t-k$. (Reason! *The sum of k and $2t-k$ is even.*) The right side of (6.34) leads to an N by N matrix $m(0)$ which is displayed for $N=5$:

$$\begin{bmatrix}\phi(0)\\ \phi(1)\\ \phi(2)\\ \phi(3)\\ \phi(4)\end{bmatrix} = 2\begin{bmatrix} h(0) & & & & \\ h(2) & h(1) & h(0) & & \\ h(4) & h(3) & h(2) & h(1) & h(0)\\ & h(5) & h(4) & h(3) & h(2)\\ & & & h(5) & h(4)\end{bmatrix}\begin{bmatrix}\phi(0)\\ \phi(1)\\ \phi(2)\\ \phi(3)\\ \phi(4)\end{bmatrix} = m(0)\Phi(0). \tag{6.35}$$

This is the dilation equation restricted to the integers. It is an eigenvalue problem for $m(0)$. That matrix has even and odd in separate columns. For a nontrivial solution, this matrix (including the factor 2) must have $\lambda = 1$ as an eigenvalue. Assume this is true. Then the values $\phi(n)$ are in the eigenvector (*which we call* $\Phi(0)$). That eigenvalue problem for $m(0)$ sets the integer values $\phi(n)$, and the recursion starts.

Now look at the vector of half-integer values. Substitute $t = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \frac{9}{2}$ into the dilation equation. This leads to a closely related matrix $m(1)$. The first row comes from $\phi(\frac{1}{2}) = 2h(1)\phi(0) + 2h(0)\phi(1)$. Notice that $2t$ is an *odd* integer, so *the sum of k and $2t-k$ is now odd.* The matrix is $m(1)$:

$$\begin{bmatrix}\phi(1/2)\\ \phi(3/2)\\ \phi(5/2)\\ \phi(7/2)\\ \phi(9/2)\end{bmatrix} = 2\begin{bmatrix} h(1) & h(0) & & & \\ h(3) & h(2) & h(1) & h(0) & \\ h(5) & h(4) & h(3) & h(2) & h(1)\\ & & h(5) & h(4) & h(3)\\ & & & & h(5)\end{bmatrix}\begin{bmatrix}\phi(0)\\ \phi(1)\\ \phi(2)\\ \phi(3)\\ \phi(4)\end{bmatrix} = m(1)\Phi(0). \tag{6.36}$$

As expected, the values at half-integers come from the values at integers. A vector $\Phi(\frac{1}{2})$ comes from a vector $\Phi(0)$. In matrix notation (6.35) was an eigenproblem for $\Phi(0)$ and (6.36) is the step to $\Phi(\frac{1}{2})$:

$$\Phi(0) = m(0)\Phi(0) \quad \text{and} \quad \Phi(\tfrac{1}{2}) = m(1)\Phi(0).$$

This is exactly like the two-coefficient case! Now $m(0)$ and $m(1)$ are $N \times N$ matrices. The beautiful fact is that the same pattern continues to quarter-integers and beyond.

When t is a quarter-integer, the times $2t-k$ are half-integers. The values $\phi(\frac{1}{4}), \phi(\frac{5}{4}), \ldots$ come from $\phi(\frac{1}{2}), \phi(\frac{3}{2}), \ldots$. The dilation equation connects those vectors by the matrix $m(0)$. Similarly the values $\phi(\frac{3}{4}), \phi(\frac{7}{4}), \ldots$ come from multiplying those half-integer values in the vector $\Phi(\frac{1}{2})$ by the matrix $m(1)$:

$$\Phi\left(\tfrac{1}{4}\right) = m(0)\,\Phi\left(\tfrac{1}{2}\right) \quad \text{and} \quad \Phi\left(\tfrac{3}{4}\right) = m(1)\Phi\left(\tfrac{1}{2}\right).$$

Exactly as before, the binary expansion of $t = n/2^i$ reveals the order of the factors $m(0)$ and $m(1)$—as they multiply the initial eigenvector $\Phi(0)$ of values at the integers. We describe the recursion and then prove it is correct.

Theorem 6.5 *The vector form of the dilation equation is*

$$\Phi(t) = m(0)\,\Phi(2t) + m(1)\,\Phi(2t-1). \tag{6.37}$$

The vector $\Phi(t)$ is zero outside the interval $0 \le t \le 1$. Its components are the N slices $\phi(t)$, $\phi(t+1)$, $\phi(t+2)$, ... of the scaling function. Substituting $t = 0$, the values $\phi(n)$ at the integer times $t = 0, 1, \ldots, N-1$ are in the eigenvector of $m(0)$ with $\lambda = 1$:

$$\text{(Fixed point)} \qquad \Phi(0) \quad = \quad m(0)\Phi(0). \tag{6.38}$$

The vector $\Phi(t)$ of values at the dyadic points $t, t+1, \ldots, t+N-1$ comes from i multiplications by $m(0)$ and $m(1)$. Here $t = n/2^i < 1$ with n odd:

$$\text{If } t = \tfrac{3}{8} = 0.011 \text{ then } \Phi(t) = \begin{bmatrix} \phi(t) \\ \phi(t+1) \\ \vdots \\ \phi(t+N-1) \end{bmatrix} = m(0)m(1)m(1)\Phi(0). \tag{6.39}$$

The scalar equation of high order is reduced to a vector equation of low order. It is just a recursion, in which the 0-1 digit t_1 tells whether to use $m(0)$ or $m(1)$:

$$\textbf{\textit{Vector recursion:}} \quad \Phi(.\,t_1\,t_2\,t_3\,\ldots) = m(t_1)\,\Phi(.\,t_2\,t_3\,t_4\,\ldots). \tag{6.40}$$

The vector $\Phi(2t)$ is nonzero on the half-interval $0 \le t < \frac{1}{2}$. The other vector $\Phi(2t-1$ is nonzero for $\frac{1}{2} \le t < 1$. These two vectors of compressed slices are multiplied by $m(0)$ an $m(1)$. To identify those particular matrices as correct, one way is to substitute t into the dilatio equation and watch the numbers $2t - k$. As t crosses $\frac{1}{2}$, those numbers cross an integer—the go from one slice to the next. At that moment $m(0)$ is exchanged for $m(1)$.

The matrix $m(0)$ has the same "*double-shift*" between rows that we saw earlier in filte banks. The earlier matrix was $L = (\downarrow 2)C$, with entries $L_{ij} = c(2i - j)$. Now this matrix appears in the dilation equation! It is multiplied by the extra factor $\sqrt{2}$ to become M. Its entrie are $2h(2i - j)$:

$$M = \sqrt{2}L \quad = \quad 2\begin{bmatrix} \cdots & h(0) & & & & \\ \cdots & h(2) & h(1) & h(0) & & \\ \cdots & h(4) & h(3) & h(2) & h(1) & h(0) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{6.4}$$

We now show that the dilation equation has an even more compact form $\Phi_\infty(t) = M\Phi_\infty(2$ These are infinite vectors and matrices. The nonzero part will be exactly the two-term form the dilation equation.

Dilation Equation in Infinite Vector Form $\Phi_\infty(t) = M\Phi_\infty(2t)$

$$\text{This form is} \quad \begin{bmatrix} \vdots \\ \phi(t-1) \\ \phi(t) \\ \phi(t+1) \\ \vdots \end{bmatrix} = M \begin{bmatrix} \vdots \\ \phi(2t-1) \\ \phi(2t) \\ \phi(2t+1) \\ \vdots \end{bmatrix} \quad \text{for } -\infty < t < \infty \tag{6.}$$

Restricted to the interval $0 \le t < 1$, only rows $0, 1, \ldots, N-1$ of $\Phi_\infty(t)$ are nonzero. Those slices of $\phi(t)$ form the vector $\Phi(t)$ with N components. The dilation equation (6.42) reduce

the vector form (6.37):

$$\Phi(t)=\left\{\begin{array}{ll} m(0)\Phi(2t) & \text{for } 0\le t<\frac{1}{2} \\ m(1)\Phi(2t-1) & \text{for } \frac{1}{2}\le t<1 \end{array}\right\} = m(0)\Phi(2t)+m(1)\Phi(2t-1).$$

The matrices $m(0)$ and $m(1)$ are N by N sections of the infinite matrix M. For $i, j = 0, 1, \ldots, N-1$ the matrix entries are

$$\boxed{m(0)_{ij} = M_{ij} = 2h(2i-j) \quad \text{and} \quad m(1)_{ij} = M_{i,j-1} = 2h(2i-j+1).}$$

Proof. The verification is in three steps, first for M and then $m(0)$ and then $m(1)$. I hope the intuition is already in place, to see the sum $\sum 2h(k)\phi(2t-k)$ as $M\Phi(2t)$ or $\sqrt{2}L\Phi(2t)$ or $2(\downarrow 2)H\Phi(2t)$. We now follow each step:

(Verify M) Row zero of $\Phi_\infty(t) = M\Phi_\infty(2t)$ is $\phi(t) = 2\sum h(k)\phi(2t-k)$.
Row one is $\phi(t+1) = 2\sum h(k)\phi(2t+2-k)$. The double shift works.

(Verify $m(0)$) Restrict to $0\le t<\frac{1}{2}$ and keep only rows $0, 1, \ldots, N-1$.
Since $\phi(2t-1)=0$ and $\phi(2t+N)=0$ we only need columns 0 to $N-1$ of the infinite matrix M. This N by N section is $m(0)$.

(Verify $m(1)$) Restrict to $\frac{1}{2}\le t<1$ and keep only rows $0, 1, \ldots, N-1$.
Since $\phi(2t-2)=0$ and $\phi(2t+N-1)=0$ we only need columns $-1, 0, \ldots, N-2$ of M. This N by N section is $m(1)$.

The change from $m(0)$ to $m(1)$ comes as t crosses $\frac{1}{2}$, because the nonzero entries in $\Phi_\infty(2t)$ appear one component earlier.

We now go back to the eigenvalue problem $\Phi(0) = m(0)\Phi(0)$. Condition A_1 leads to the eigenvalue $\lambda = 1$, and guarantees a solution.

The Fixed Point Equation $\Phi(0) = m(0)\Phi(0)$

The rows of $m(0)$ have a double shift. The columns have entirely even indices or entirely odd indices. The 5 by 5 matrix ($N=5$) shows this pattern:

$$m(0) \quad = \quad 2\begin{bmatrix} h(0) & & & & \\ h(2) & h(1) & h(0) & & \\ h(4) & h(3) & h(2) & h(1) & h(0) \\ & h(5) & h(4) & h(3) & h(2) \\ & & & h(5) & h(4) \end{bmatrix}.$$

The key requirement on the coefficients $h(n)$ is Condition A_1: *The frequency response $H(\omega)$ has a zero at $\omega=\pi$:*

$$h(0)-h(1)+h(2)-\cdots=0.$$

Combined with $h(0)+h(1)+h(2)+\cdots=1$, this means that every column of $m(0)$ and $m(1)$ and M adds to 1.

Theorem 6.6 *Condition A_1 guarantees that $\lambda = 1$ is an eigenvalue of M and $m(0)$ and $m(1)$:*

$$\begin{array}{lcl} h(0)-h(1)+h(2)-\cdots & = & 0 \\ h(0)+h(1)+h(2)+\cdots & = & 1 \end{array} \quad \textit{yields} \quad 2\sum_{\text{even } n} h(n) = 2\sum_{\text{odd } n} h(n) = 1.$$

Any matrix with unit column sums has $\lambda = 1$ as an eigenvalue. Therefore $\Phi(0) = m(0)\Phi(0)$ can be solved to give the scaling function at integer times $\Phi(0) = (\phi(0), \phi(1), \ldots, \phi(N-1))^T$.

Proof. Adding the two equations gives the even part. Subtracting gives the odd part. These are the column sums of the matrix $m(0)$, all equal to 1. A matrix with unit column sums has a *left* eigenvector of ones, $em(0) = e$, because the multiplication just adds up each column:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} m(0) = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}.$$

This means that $m(0) - I$ is not invertible, and $\lambda = 1$ is an eigenvalue. The left eigenvector is e. The right eigenvector is $(\phi(0), \phi(1), \ldots, \phi(N-1))$:

$$(m(0) - I)\Phi(0) = 0 \text{ which means } \Phi(0) = m(0)\Phi(0).$$

The fundamental fact is that a square matrix and its transpose have the same determinant and same rank and same eigenvalues.

The columns of $m(1)$ and M also add to 1, producing the eigenvalue $\lambda = 1$. Small point! We can safely normalize the eigenvector $\Phi(0)$ by $\sum \phi(n) = 1$. This is the "unit area" requirement that we impose on the function $\phi^{(0)}(t)$ at the start of the iterations. Then the scaling function has $\int \phi(t)dt = 1$ at the end.

Corollary *The sum* $\sum \phi(t+k)$ *is identically* 1.

Proof. Multiply the vector dilation equation $\Phi(t) = m(0)\,\Phi(2t) + m(1)\,\Phi(2t-1)$ on the left by e. Use the fact that $e\,m(0) = e$ and $e\,m(1) = e$:

$$e\,\Phi(t) = e\,\Phi(2t) + e\,\Phi(2t-1).$$

This is a dilation equation for $e\,\Phi(t)$ and its solution is the box function! Thus $e\,\Phi(t) = 1$. The "periodized scaling function" $\sum \phi(t+k)$ is identically one.

Example 6.3. The coefficients $2h(k) = \frac{1}{2}, 1, \frac{1}{2}$ lead to the hat function. The 2 by 2 eigenvalue problem for $m(0)$ gives the correct values of $\phi(0)$ and $\phi(1)$:

$$\begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \phi(0) \\ \phi(1) \end{bmatrix} = \begin{bmatrix} \phi(0) \\ \phi(1) \end{bmatrix} \quad \text{gives} \quad \begin{matrix} \phi(0) = 0 \\ \phi(1) = 1. \end{matrix}$$

The sum of all hat functions $\phi(t+n)$ is identically one. Notice that the first row of the eigenvalue equation is always $2h(0)\phi(0) = \phi(0)$. Then $\phi(0)$ is zero, apart from the exceptional case $h(0) = \frac{1}{2}$ which occurs for the box function. This means that the scaling function $\phi(t)$ is zero up to and *including* $t = 0$. The box function starts with a jump at $t = 0$, because $h(0) = \frac{1}{2}$.

Example 6.4. The Daubechies coefficients have $8h(k) = 1+\sqrt{3}, 3+\sqrt{3}, 3-\sqrt{3}$, and $1-\sqrt{3}$. Dividing by 4 we have $2h(k)$, the numbers that enter $m(0)$:

$$\frac{1}{4}\begin{bmatrix} 1+\sqrt{3} & 0 & 0 \\ 3-\sqrt{3} & 3+\sqrt{3} & 1+\sqrt{3} \\ 0 & 1-\sqrt{3} & 3-\sqrt{3} \end{bmatrix} \begin{bmatrix} \phi(0) \\ \phi(1) \\ \phi(2) \end{bmatrix} = \begin{bmatrix} \phi(0) \\ \phi(1) \\ \phi(2) \end{bmatrix}.$$

The eigenvector gives $\phi(0)$, $\phi(1)$, and $\phi(2)$:

$$\phi(0) = 0 \qquad \phi(1) = \tfrac{1}{2}\left(1+\sqrt{3}\right) \qquad \phi(2) = \tfrac{1}{2}\left(1-\sqrt{3}\right).$$

We now know the Daubechies scaling function $\phi(t)$ at the integers. The only nonzeros in the fixed-point eigenvector Φ are $\phi(1)$ and $\phi(2)$. From these values at $t = 1$ and $t = 2$, the recursion produces $\phi(t)$ at any dyadic point.

Practical conclusion Every $\phi\left(n/2^i\right)$ comes via $\boldsymbol{m}(0)$ and $\boldsymbol{m}(1)$.

Theoretical conclusion Those dyadic values have a uniform bound if and only if all products of $\boldsymbol{m}(0)$ and $\boldsymbol{m}(1)$ in all orders have a *uniform upper bound.* When this holds, the dilation equation has a bounded solution $\phi(t)$ for all t.

We can propose sufficient conditions so that all products of $\boldsymbol{m}(0)$ and $\boldsymbol{m}(1)$ have a uniform bound. We can also propose necessary conditions. It is not known how to verify the necessary and sufficient condition (Section 7.3).

Derivatives of the Dilation Equation

While working in the time domain, we might as well take the derivative of $\phi(t)$. The result is highly interesting and not fully understood. Part of the problem is that the derivative $\phi'(t)$ may not exist.

The plan is to differentiate each term in the dilation equation for $\phi(t)$:

$$\phi'(t) \quad = \quad 4\sum \boldsymbol{h}(k)\phi'(2t-k).$$

This is another dilation equation, with every coefficient doubled. The equation $\Phi(t) = M\Phi(2t)$ has led to $\Phi'(t) = 2M\Phi'(2t)$. At $t = 0$ this yields the fixed-point equation $\Phi'(0) = 2M\Phi'(0)$. The eigenvector $\Phi'(0)$ contains the derivatives $\phi'(n)$ at the integers $t = n$.

To solve $\Phi'(0) = M\Phi'(0)$, we have a new requirement. *The number $\lambda = \frac{1}{2}$ must also be an eigenvalue of M.* Again this applies to the $N \times N$ matrices $\boldsymbol{m}(0)$ and $\boldsymbol{m}(1)$. This new requirement on the entries is stated as Condition A_2 in the following theorem.

Theorem 6.7 *The matrices M and $\boldsymbol{m}(0)$ and $\boldsymbol{m}(1)$ have eigenvalues 1 and $\frac{1}{2}$ if and only if the filter coefficients satisfy Condition A_2 which includes A_1:*

$$\textit{Condition } A_2: \quad \sum_0^N (-1)^k \boldsymbol{h}(k) = 0 \quad \textit{and} \quad \sum_0^N (-1)^k k\boldsymbol{h}(k) = 0.$$

The eigenvector for $\lambda = 1$ is $\Phi(0)$, containing the values $\phi(0), \ldots, \phi(N-1)$. The eigenvector for $\lambda = \frac{1}{2}$ is $\Phi'(0)$, containing the derivatives $\phi'(0), \ldots, \phi'(N-1)$.

In this case $H(\omega)$ has a *double zero* at $\omega = \pi$. This beautiful pattern extends onward to Condition A_p. The matrices have eigenvalues $\lambda = 1, \frac{1}{2}, \ldots, \left(\frac{1}{2}\right)^{p-1}$ if and only if the filter coefficients satisfy p sum rules:

$$\textit{Condition } A_p: \quad \sum (-1)^k k^m \boldsymbol{h}(k) = 0 \quad \text{for } m = 0, \ldots, p-1.$$

The eigenvector for $\lambda = \left(\frac{1}{2}\right)^m$ contains values of the m^{th} derivative of $\phi(t)$ at the integers. Formally, $\Phi^{(m)}(0) = 2^m M \Phi^{(m)}(0)$ comes from differentiating the dilation equation m times at the integers. We mention the frequency domain equivalent:

Condition A_p: The frequency response $H(\omega)$ has a zero of order p at $\omega = \pi$.

We will see this again! And we also begin to uncover the crucial role of the *left eigenvectors* (which are row vectors). Those tell how to produce polynomials from combinations of the translates $\phi(t-k)$. Under Condition A_p, these low order polynomials $1, t, \ldots, t^{p-1}$ are in the lowpass space V_0. They are the keys to approximation of a function $f(t)$ by functions in V_0.

The letters A_p *indicate "approximation of order p."* The theorem above, with its extension from 2 to p, is absolutely basic to the algebra of downsampled filters. These eigenvalues and eigenvectors control everything in Chapter 7.

You will see that the derivative $\phi'(t)$ is often *one-sided.* Derivatives of $\phi(t)$ may not exist in the usual sense. This subject still contains some mysteries.

Example 6.5. The hat function coefficients $2h(k) = \frac{1}{2}, 1, \frac{1}{2}$ satisfy Condition A_2:

$$\begin{aligned} &\text{First sum rule:} && \tfrac{1}{2} - 1 + \tfrac{1}{2} = 0 \\ &\text{Second sum rule:} && 0\left(\tfrac{1}{2}\right) - 1(1) + 2\left(\tfrac{1}{2}\right) = 0. \end{aligned}$$

Therefore $m(0)$ will have eigenvalues 1 and $\frac{1}{2}$:

$$m(0) = \begin{bmatrix} 2h(0) & 0 \\ 2h(2) & 2h(1) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{bmatrix}.$$

The eigenvector for $\lambda = 1$ has components 0 and 1. They agree with $\phi(0)$ and $\phi(1)$, the hat function at the integers. The eigenvector for $\lambda = \frac{1}{2}$ has components 1 and -1. They are $\phi'_+(0)$ and $\phi'_+(1)$, the slopes $\phi'(t)$ of the hat function in the two intervals. These are derivatives *from the right* at the points $t = 0$ and $t = 1$. The slopes on the left side of those points are different because the hat function has corners.

The matrix $m(1)$ must also have eigenvalues 1 and $\frac{1}{2}$:

$$m(1) = \begin{bmatrix} 2h(1) & 2h(0) \\ & 2h(2) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}.$$

The eigenvector for $\lambda = 1$ has components 1 and 0. Those agree with the hat function at the shifted points $t = 1$ and $t = 2$. The eigenvector for $\lambda = \frac{1}{2}$ has components 1 and -1. Those agree with the slopes of the hat function *from the left* at $t = 1$ and $t = 2$. Remember that $m(0)$ is involved at the start of an interval and $m(1)$ is involved at the end of an interval.

Condition A_3 is not satisfied for the hat function. There is no eigenvalue $\lambda = \frac{1}{4}$. The hat function has no second derivatives at the integers.

Problem Set 6.3

1. If the filter $H(z)$ is halfband, show that the eigenvector in $m(0)\Phi = \Phi$ is an impulse $\delta(n)$. What are the values of $\phi(t)$ at the integers?

. If $\phi_1(t)$ and $\phi_2(t)$ satisfy dilation equations, does their product $P(t) = \phi_1(t)\phi_2(t)$ satisfy a dilation equation?

. Show that the convolution $\phi_1(t) * \phi_2(t)$ *does* satisfy a dilation equation with coefficients from $h_1 * h_2$.

. Find a specific function $f(t)$ that does not satisfy any dilation equation.

Infinite Product Formula

caling function $\phi(t)$ comes from the dilation equation

$$\phi(t) \quad = \quad 2\sum_{k=0}^{N} h(k)\phi(2t-k).$$

$\phi(t)$ is the fixed point, or fixed function, when we iterate with $\boldsymbol{H}$ and rescale. In the time in, the matrix that filters and rescales is $\boldsymbol{M} = (\downarrow 2)2\boldsymbol{H}$. Now we intend to find the Fourier orm $\widehat{\phi}(\omega)$ in the frequency domain. Just as the time-domain solution involved products of and $\boldsymbol{m}(1)$, the frequency-domain solution will involve an infinite product of $H(\omega)$'s.
is quite remarkable that two-scale equations received so little attention for so long. His- lly, t and $2t$ were not often seen in the same equation. They began to appear prominently ıctals, which are self-similar. Now, multiple scales seem to be everywhere. We meet them ; book through multirate filters—with two scales. Then the iteration leads to all scales.
the 2's were removed, the dilation equation would be an ordinary difference equation. oefficients are constant, so we look for pure exponential solutions $e^{i\omega t}$. When you make ıbstitution, you are effectively taking the Fourier transform of the equation. The trans- urns difference equations and differential equations (*and dilation equations*) into algebraic ons. We do that now for the two-scale equation, and we watch how $2t$ leads to $\omega/2$.
ıe dilation equation becomes $\widehat{\phi}(\omega) = H(\frac{\omega}{2})\,\widehat{\phi}(\frac{\omega}{2})$. This leads recursively to an infinite ct for $\widehat{\phi}(\omega)$. This transform must be a *sinc function* when $h(0) = h(1) = 1/2$ —because ne-domain solution $\phi(t)$ is a box function. That sinc function must be orthogonal to its ations by $e^{-i\omega k}$, because the box function is orthogonal to its translates $\phi(t-k)$. We ɔ study orthogonality and also approximation, which is controlled by "zeros at π." These ties are now studied in the frequency domain.

Condition **O** for orthogonality:

$$\boxed{\begin{array}{lcl} |H(\omega)|^2 + |H(\omega+\pi)|^2 & \equiv & 1 \text{ in the frequency domain} \\ 2\sum h(k)h(k-2\ell) & = & \delta(\ell) \text{ in the time domain} \end{array}}$$

double-shift orthogonality of the lowpass filter coefficients. It connects to orthogonality scaling functions $\phi(t-k)$. We use the word "connects" rather than "implies," because a condition is eventually needed to insure orthogonality in the limit of the iteration. The ɔm discrete time to continuous time seldom goes wrong, but it can.
is orthogonality will appear in Theorem 6.10 as a neat statement about the Fourier trans- ɔf ϕ and w:

$$\sum_{-\infty}^{\infty} |\widehat{\phi}(\omega+2\pi n)|^2 = 1 \quad \text{and} \quad \sum_{-\infty}^{\infty} \widehat{\phi}(\omega+2\pi n)\overline{\widehat{w}(\omega+2\pi n)} = 0.$$

The other side of the theory is about approximation. This imposes a very different condition on the $h(k)$ and the polynomial $H(\omega) = \sum h(k)e^{-ik\omega}$.

2. Condition A_p: $H(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^p Q(\omega)$.

This factor $\left(1+e^{-i\omega}\right)^p$ means that $H(\omega)$ has a *zero of order* p at $\omega = \pi$. We will prove tha this puts the polynomials $1, t, \ldots, t^{p-1}$ in the scaling subspace V_0. They are combinations o $\phi(t-n)$ and they are orthogonal to $w(t-n)$. In the frequency domain, there is again a nea statement about the Fourier transforms of ϕ and w:

$$\widehat{\phi} \text{ has a zero of order } p \text{ at every } \omega = 2\pi n, n \neq 0$$
$$\widehat{w} \text{ has a zero of order } p \text{ at zero frequency.}$$

The wavelet coefficients of a smooth function $f(t) = \sum b_{jk}w_{jk}(t)$ decrease faster whe p is larger. The estimate is $|b_{jk}| = O\left(2^{-jp}\right)$. This is valuable for compression. This sectio does the frequency-domain algebra, to solve the dilation equation and to explain Condition (and Condition A_p.

Transform and Solution of the Dilation Equation

To transform the dilation equation, multiply by $e^{-i\omega t}$. Integrate with respect to t:

$$\int_{-\infty}^{\infty} \phi(t)e^{-i\omega t}\,dt \quad = \quad 2\sum_{k=0}^{N} h(k)\int_{-\infty}^{\infty} \phi(2t-k)e^{-i\omega t}\,dt.$$

The left side is $\widehat{\phi}(\omega)$. In the integral on the right, set $u = 2t-k$ and $t = (u+k)/2$:

$$\int_{-\infty}^{\infty} \phi(2t-k)e^{-i\omega t}2\,dt = \int_{-\infty}^{\infty} \phi(u)e^{-i\omega(u+k)/2}\,du = e^{-i\omega k/2}\,\widehat{\phi}\left(\tfrac{\omega}{2}\right). \qquad (6.4$$

Instead of t and $2t$, *the transform involves* ω *and* $\omega/2$. The dilation equation becomes

$$\widehat{\phi}(\omega) = \left(\sum h(k)e^{-i\omega k/2}\right)\widehat{\phi}\left(\tfrac{\omega}{2}\right) = H\left(\tfrac{\omega}{2}\right)\widehat{\phi}\left(\tfrac{\omega}{2}\right). \qquad (6.4$$

This is the result of filtering and rescaling. Filtering multiplies $\widehat{\phi}(\omega)$ by $H(\omega)$. Rescali changes ω to $\omega/2$. The scaling function (which is unique up to a constant multiple C — tl is still to be proved) comes out unchanged:

$$\boxed{\widehat{\phi}(\omega) = H\left(\tfrac{\omega}{2}\right)\widehat{\phi}\left(\tfrac{\omega}{2}\right).}$$

Now iterate this equation. It connects ω to $\omega/2$ and therefore it connects $\omega/2$ to $\omega/4$:

$$\widehat{\phi}(\omega) \quad = \quad H\left(\tfrac{\omega}{2}\right)\left[H\left(\tfrac{\omega}{4}\right)\widehat{\phi}\left(\tfrac{\omega}{4}\right)\right].$$

After N iterations, this becomes

$$\widehat{\phi}(\omega) = H\left(\tfrac{\omega}{2}\right)H\left(\tfrac{\omega}{4}\right)\cdots H\left(\tfrac{\omega}{2^N}\right)\widehat{\phi}\left(\tfrac{\omega}{2^N}\right).$$

In the limit as $N \to \infty$, we have a formula for the solution $\widehat{\phi}(\omega)$. Note that $\omega/2^N$ is approach zero, and $\widehat{\phi}(0) = \int \phi(t)\,dt$ is the area under the graph of $\phi(t)$. *This equals one.* We imp

the normalization $\widehat{\phi}(0) = 1$ in the frequency domain, just as we required unit area in the time domain. Then the formal limit of the iteration leads to the famous infinite product for $\widehat{\phi}$:

$$\boxed{\widehat{\phi}(\omega) = \prod_{j=1}^{\infty} H\left(\tfrac{\omega}{2^j}\right).} \tag{6.45}$$

We note the minimum requirement for convergence of this or any infinite product: the factors $H(\omega/2^j)$ must approach 1 as $j \to \infty$. Thus we need $H(0) = 1$. By periodicity $H(2\pi) = H(4\pi) = 1$. Then the equation $\widehat{\phi}(\omega) = H\left(\frac{\omega}{2}\right)\widehat{\phi}\left(\frac{\omega}{2}\right)$ has a remarkable consequence. *The values $\widehat{\phi}(2\pi)$, $\widehat{\phi}(4\pi)$, $\widehat{\phi}(8\pi)$, ... are all equal. If $H(\pi) = 0$, those values equal zero because $\widehat{\phi}(2\pi) = H(\pi)\,\widehat{\phi}(\pi)$.*

This "zero at π" is a natural requirement on $H(\omega)$ in order that $\widehat{\phi}(\omega)$ may decay and $\phi(t)$ may be a reasonable function.

The infinite product converges for every ω and every $H(\omega)$. We have an explicit formula for $\widehat{\phi}(\omega)$. Whether any function $\phi(t)$ has this Fourier transform is another matter! Convergence follows from a rough bound on $H(\omega)$ in terms of $C = \max |H'(\omega)|$:

$$|H(\omega)| = |1 + H(\omega) - H(0)| \le 1 + C|\omega| \le e^{C|\omega|}.$$

Then the product $\widehat{\phi}(\omega)$ has the same upper bound:

$$|\widehat{\phi}(\omega)| = \left|H\left(\tfrac{\omega}{2}\right)\right| \left|H\left(\tfrac{\omega}{4}\right)\right| \cdots \le e^{C|\omega|/2} e^{C|\omega|/4} \cdots = e^{C|\omega|}.$$

This is a wild overestimate of $\widehat{\phi}(\omega)$, as almost any example will show.

Box example. The coefficients are $h(0) = h(1) = \frac{1}{2}$. Then $H(\omega) = \frac{1}{2}\left(1 + e^{-i\omega}\right)$. The product of the first N factors contains 2^N terms. Looked at correctly, those terms are the first 2^N powers of $e^{-i\omega/2^N}$:

$$\begin{aligned} H^{(N)}(\omega) &= \frac{1}{2^N}\left(1 + e^{-i\omega/2}\right)\left(1 + e^{-i\omega/4}\right)\cdots\left(1 + e^{-i\omega/2^N}\right) \\ &= \frac{1}{2^N}\sum_{k=0}^{2^N-1} e^{-i\omega k/2^N} \quad \text{(geometric series)} \\ &= \frac{1 - e^{-i\omega}}{2^N\left(1 - e^{-i\omega/2^N}\right)} \quad \text{(sum of } 2^N \text{ terms).} \end{aligned} \tag{6.46}$$

Now let $N \to \infty$. The denominator has $1 - e^{-i\theta} = 1 - (1 - i\theta + \cdots) = i\theta + \cdots$ with $\theta = \omega/2^N$. The limit of $2^N i\theta$ is $i\omega$. Therefore the limit of the partial product is the infinite product

$$\widehat{\phi}(\omega) \quad = \quad \prod\left(\tfrac{1}{2} + \tfrac{1}{2}e^{-i\omega/2^j}\right) = (1 - e^{-i\omega})/i\omega. \tag{6.47}$$

This sinc function is the transform of the box function. The integral of $e^{-i\omega t}$ from 0 to 1 agrees with $\widehat{\phi}(\omega)$. Instead of increasing like $e^{C|\omega|}$, as allowed by the general estimate, the transform $\widehat{\phi}(\omega)$ actually decreases to zero as ω becomes large.

Compare the construction of $\phi(t)$ with $\widehat{\phi}(\omega)$. In Section 6.2, we assumed that $\phi^{(i)}(t)$ converged uniformly to $\phi(t)$. Then we studied its properties. In this section, the convergence of the infinite product is cheap (for each separate ω). What we need is sufficient decay of $\widehat{\phi}(\omega)$ as

$|\omega| \to \infty$. Our precise assumption will be continuity of the function $A(\omega)$ in Theorem 6.10 below. Then we can safely study $\widehat{\phi}(\omega)$ in the frequency domain. Note first that for real frequencies, the growth of $\widehat{\phi}(\omega)$ is at most polynomial:

Theorem 6.8 $|\widehat{\phi}(\omega)| \le e^{C|\omega|}$ *for complex* ω *and* $|\widehat{\phi}(\omega)| \le c(1+|\omega|^M)$ *for real* ω.

Brief reason: $H(\omega)$ is periodic. It has a maximum value 2^M. The equation $\widehat{\phi}(2\omega) = H(\omega)\widehat{\phi}(\omega)$ says that $\widehat{\phi}$ grows by at most 2^M when ω is doubled. The bound $|\omega|^M$ has this growth rate. A constant is included to make $c(1+|\omega|^M)$ correct for small $|\omega|$.

The example $H(\omega) = \frac{1}{2} + \frac{1}{2}e^{-i\omega}$ is bounded by 1 for real ω and by $e^{|\omega|}$ for complex ω. Then $M = 0$. The transform $\widehat{\phi}(\omega)$ of the box function has those same bounds:

$$|\widehat{\phi}(\omega)| = \left|H(\tfrac{\omega}{2})\right| \left|H(\tfrac{\omega}{4})\right| \cdots \le \begin{cases} 1 & \text{for real } \omega \\ e^{|\omega|/2}\, e^{|\omega|/4} \cdots = e^{|\omega|} & \text{for complex } \omega. \end{cases}$$

Section 6.1 showed that the support interval for $\phi(t)$ is $[0, N]$. This can be proved in the frequency domain too. Our bounds on $\widehat{\phi}(\omega)$ show two fundamental facts about $\phi(t)$:

Theorem 6.9 *Any dilation equation with* $h(0)+\cdots+h(N) = 1$ *has a* unique and compactly supported solution $\phi(t)$. *This solution may be a distribution.*

Compact support comes from $|\widehat{\phi}(\omega)| \le e^{C|\omega|}$. The Paley-Wiener Theorem implies that $\phi(t)$ is supported on the interval $[-C, C]$. With more care [D, p.176] we could find again the exact support interval $[0, N]$.

Uniqueness comes from our formula for the solution! The infinite product converges to $\widehat{\phi}(\omega)$, which is continuous because $\phi(t)$ has compact support:

$$\widehat{\phi}^{(i)}(\omega) = \Big(\prod_{j=1}^{i} H(\omega/2^j)\Big)\,\widehat{\phi}(\omega/2^i) \text{ approaches } \widehat{\phi}(\omega) = \Big(\prod_{1}^{\infty} H(\omega/2^j)\Big)\,\widehat{\phi}(0).$$

In the IIR case, suitable hypotheses will again give uniqueness (of course not compact support). At the other extreme, note how the lazy filter with $h(0) = 1$ leads to $\phi(t) =$ delta function. *The dilation equation* $\phi(t) = 2\phi(2t)$ *is solved by* $\phi(t) = \delta(t)$:

In frequency: $H(\omega) \equiv 1$ so $\widehat{\phi}(\omega) = 1$.
Cascade algorithm: $\phi^{(i)}(t) =$ box function on $[0, 2^{-i}]$ with height 2^i.
Verify directly: $\delta(t) = 2\delta(2t)$ from $\int f(t)\delta(t)dt = f(0) = \int f(t)\delta(2t)2\,dt$.

All these methods show that $h(0) = 1$ produces the best-known distribution $\phi(t) = \delta(t)$.

Orthogonality in the Frequency Domain

The product formula for $\widehat{\phi}(\omega)$ applies with or without Condition O. When that condition holds, we expect orthogonality of the translates $\phi(t-k)$. To establish this orthogonality in the frequency domain, we need to know that the equivalent statement is $A(\omega) \equiv 1$. The function $A(\omega)$ enters naturally into this discussion. It is the transform $\sum a(k)e^{i\omega k}$ of the vector of *inner products* of $\phi(t)$ with $\phi(t-k)$:

Theorem 6.10 *The inner products $a(k)$ are the Fourier coefficients of the 2π-periodic function $A(\omega)$:*

$$a(k) = \int_{-\infty}^{\infty} \phi(t)\overline{\phi(t-k)}\,dt \quad \textit{transforms to} \quad A(\omega) = \sum_{-\infty}^{\infty} \left|\widehat{\phi}(\omega + 2\pi n)\right|^2.$$

The translates $\phi(t-k)$ are orthonormal if and only if $A(\omega) \equiv 1$.

Proof. An inner product in the time domain equals an inner product in the frequency domain, by Parseval's identity. The inner product in the time domain is between $\phi(t)$ and $\phi(t-k)$. The transforms of these functions are $\widehat{\phi}(\omega)$ and $e^{-i\omega k}\widehat{\phi}(\omega)$. Each inner product integrates one function times the complex conjugate of the other:

$$\begin{aligned} a(k) = \int_{-\infty}^{\infty} \phi(t)\overline{\phi(t-k)}\,dt &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\phi}(\omega)\overline{\widehat{\phi}(\omega)}e^{i\omega k}\,d\omega \\ &= \frac{1}{2\pi}\int_{0}^{2\pi} \sum_{-\infty}^{\infty} \left|\widehat{\phi}(\omega + 2\pi n)\right|^2 e^{i\omega k}\,d\omega. \end{aligned} \tag{6.48}$$

The last integral split $(-\infty, \infty)$ into an infinite number of 2π-pieces, using the periodicity of $e^{i\omega k}$. This integral defines the k^{th} Fourier coefficient of $A(\omega)$. Thus $A(\omega) = \sum a(k)e^{i\omega k}$.

For an FIR filter, $\phi(t) = 0$ outside the interval $[0, N]$. The inner products are $a(k) = 0$ for $|k| > N$, because $\phi(t)$ and $\phi(t-k)$ have no overlap. *The function $A(\omega) = \sum a(k)e^{i\omega k}$ is a trigonometric polynomial of degree N*, which is not obvious from $\sum |\widehat{\phi}(\omega + 2\pi n)|^2$. In Section 7.3 we will compute $a(k)$ directly from the coefficients $h(n)$. This is always a main point of the theory, to return every calculation to those numbers $h(n)$.

When the translates are orthonormal, all inner products $a(k)$ are zero except for $a(0) = 1$. The function with those coefficients is the constant function $A(\omega) \equiv 1$:

$$\boxed{\phi(t-k) \quad \text{are orthonormal} \quad \Longleftrightarrow \quad \sum_{-\infty}^{\infty} \left|\widehat{\phi}(\omega + 2\pi n)\right|^2 \equiv 1.}$$

We now apply Condition O in the frequency domain to deduce this orthogonality of $\phi(t-k)$. We are repeating in the frequency domain the result of Section 6.2 in the time domain. I believe this is worthwhile! The arguments in the two domains look quite different. Recall the condition on the frequency response $H(\omega)$ to produce an orthonormal filter bank:

$$\text{Condition O:} \quad |H(\omega)|^2 + |H(\omega + \pi)|^2 \equiv 1.$$

This function $H(\omega)$ leads to $\widehat{\phi}(\omega)$ which leads to $A(\omega)$. Somehow, Condition O must imply that $A(\omega) \equiv 1$. The steps are typical of computations in the frequency domain.

Theorem 6.11 *If $A(\omega)$ is continuous, $A(\omega) \equiv 1$ is equivalent to Condition O.*

Proof. We use a very important two-scale identity, proved below:

$$A(2\omega) = |H(\omega)|^2 A(\omega) + |H(\omega + \pi)|^2 A(\omega + \pi). \tag{6.49}$$

If $A(\omega) \equiv 1$, this immediately gives that $|H(\omega)|^2 + |H(\omega + \pi)|^2 \equiv 1$.

For the converse, suppose that Condition O holds. The identity says that $A(2\omega)$ is a weighted average of $A(\omega)$ and $A(\omega+\pi)$. At the point ω_0 where $A(2\omega)$ reaches its maximum, $A(\omega_0)$ and $A(\omega_0+\pi)$ must also reach that maximum. Now repeat the argument at $\omega_0/2$, to show that $A(\omega_0/2)$ shares this same maximum with $A(\omega_0)$. Continuing, the maximum of $A(\omega)$ is achieved at $\omega_0/4$ and $\omega_0/8$ and eventually (by continuity) at $\omega=0$.

By a similar argument, which is due to Tchamitchian, the minimum of $A(\omega)$ is also attained at $\omega=0$. *Therefore $A(\omega)$ is constant.* We verify below that the constant is one: $A(\omega)\equiv 1$. It only remains to prove (6.49).

This valuable identity for $A(2\omega)$ has a nice proof. It uses the dilation equation $\widehat{\phi}(2\omega) = H(\omega)\widehat{\phi}(\omega)$. At the points $2\omega+2\pi n$, this splits into separate cases for even n and odd n:

$$\begin{aligned}\widehat{\phi}(2\omega+2\pi n) &= H(\omega+\pi n)\,\widehat{\phi}(\omega+\pi n)\\ &= \begin{cases} H(\omega)\,\widehat{\phi}(\omega+2k\pi) & \text{if } n=2k\\ H(\omega+\pi)\,\widehat{\phi}(\omega+(2k+1)\pi) & \text{if } n=2k+1.\end{cases}\end{aligned}$$

Now square both sides. Sum from $-\infty$ to ∞ on n and therefore on k. The sum of squares is our function $A(2\omega)$ in the desired identity:

$$\begin{aligned}A(2\omega) = |H(\omega)|^2\sum_{-\infty}^{\infty}\left|\widehat{\phi}(\omega+2\pi k)\right|^2 + |H(\omega+\pi)|^2\sum_{-\infty}^{\infty}\left|\widehat{\phi}(\omega+\pi+2\pi k)\right|^2\\ = |H(\omega)|^2 A(\omega) + |H(\omega+\pi)|^2 A(\omega+\pi).\end{aligned} \tag{6.50}$$

The final step is to confirm that $A(0)=1$. This comes from our other condition on the lowpass filter, not yet used in the frequency domain. Condition A_1 is $H(\pi)=0$. In the time domain, this first sum rule guaranteed an eigenvalue $\lambda=1$ for the matrices $\boldsymbol{M}$ and $\boldsymbol{m}(0)$ and $\boldsymbol{m}(1)$. The fixed-point equation $\phi^{(1)}(n)=\phi^{(0)}(n)$ at the integers could be solved. Condition A_1 is equally essential in the frequency domain. Here we use it to pin down the value $A(0)=1$.

Theorem 6.12 *If $H(\pi)=0$ then $\widehat{\phi}(2\pi n)=0$ for all $n\neq 0$. Therefore*

$$A(0)=\sum_{-\infty}^{\infty}\left|\widehat{\phi}(2\pi n)\right|^2 = \left|\widehat{\phi}(0)\right|^2 = 1. \tag{6.51}$$

Proof. The infinite product for $\widehat{\phi}(2\pi)=H(\pi)H(\pi/2)\cdots$ starts with the factor $H(\pi)$. Immediately this product is zero. For any higher value $n>1$, write $n=2^j m$ with odd m. Then the $(j+1)^{\text{st}}$ factor in the infinite product is zero when $\omega=2\pi n$:

$$\widehat{\phi}(2\pi n)=H(\pi n)H\left(\frac{\pi n}{2}\right)\cdots H\left(\frac{\pi n}{2^j}\right)\cdots=0$$

because $H(\pi n/2^j) = H(\pi m)$. By periodicity this is $H(\pi)=0$. The only nonzero term is $|\widehat{\phi}(0)|^2$. But $\widehat{\phi}(0)=H(0)H(0)H(0)\cdots$ which is 1.

Orthogonalization of the Basis

The condition for an orthonormal basis is $A(\omega)\equiv 1$. When this is not satisfied, there is an easy way to *make* it satisfied. In other words: when the translates $\phi(t-n)$ are not orthonormal, there

is an easy way to *make* them orthonormal. Divide $\widehat{\phi}(\omega)$ by the given $A(\omega)$ (or rather, its square root) to get the new orthogonalized function $\widehat{\phi}_{\text{orth}}(\omega)$:

$$\widehat{\phi}_{\text{orth}}(\omega) = \frac{\widehat{\phi}(\omega)}{\sqrt{A(\omega)}}.$$

This immediately gives orthogonality of the new basis $\{\phi_{\text{orth}}(t-n)\}$:

$$A_{\text{orth}}(\omega) = \sum \left| \frac{\widehat{\phi}(\omega + 2\pi n)}{\sqrt{A(\omega)}} \right|^2 = \frac{A(\omega)}{A(\omega)} \equiv 1.$$

That succeeds if $A(\omega)$ is never zero. This is the condition for a ***Riesz basis***.

Theorem 6.13 *The upper and lower bounds on $A(\omega)$ are the Riesz constants B and A for the basis $\{\phi(t-k)\}$ of V_0. Thus $A(\omega) \geq A > 0$ gives a stable basis, and dividing $\widehat{\phi}(\omega)$ by $\sqrt{A(\omega)}$ gives an orthonormal basis.*

When $\phi(t)$ comes from a dilation equation — this is our normal situation — Condition **E** in Chapter 7 gives an equivalent test for a Riesz basis (in terms of eigenvalues). If this test is passed, the wavelets $w_{jk}(t)$ are a Riesz basis for $L^2(\mathbf{R})$.

Proof. To test the linear independence of the functions $\phi(t-k)$, form the matrix A from their inner products. The entries are $A_{ij} = \langle \phi(t-i), \phi(t-j) \rangle$. That number is $a(j-i)$, and A is a Toeplitz matrix! It is the matrix TT^* in Section 2.5. In the frequency domain it becomes multiplication by $A(\omega)$. The upper and lower bounds on $A(\omega)$ determine whether $\{\phi(t-k)\}$ is a Riesz basis.

Orthogonalization is always a basic step in linear algebra. There it is done by the Gram-Schmidt algorithm. We start with independent vectors and produce orthonormal vectors (or functions). This algorithm is not successful here, because it is not time-invariant. The orthogonalized functions will certainly not be translates — when the Gram-Schmidt algorithm works on functions in a definite order like $\phi(t), \phi(t-1), \phi(t+1), \ldots$. To keep a shift-invariant basis, we needed to orthogonalize all these translates at once. The division by $\sqrt{A(\omega)}$ did it.

In matrix language, $\boldsymbol{M}_{\text{orth}}\boldsymbol{M}_{\text{orth}}^T = \boldsymbol{I}$. In the improved factorization by Fourier methods, all rows of $\boldsymbol{M}_{\text{orth}}$ come from the zeroth row by double shifts. In other words, $\boldsymbol{M}_{\text{orth}}$ comes from a filter.

One problem with dividing by $\sqrt{A(\omega)}$. This destroys the finite response of the original filter $\boldsymbol{H}$. The new filter $\boldsymbol{H}_{\text{orth}}$ is IIR, not FIR. The new scaling function $\phi_{\text{orth}}(t)$ that corresponds to $\widehat{\phi}(\omega)/\sqrt{A(\omega)}$ does not have compact support. Vetterli and Herley noticed that this is not as bad as it seems. Since $\phi(t)$ is zero outside the interval $[0, N]$, the inner products $a(k) = \int \phi(t)\overline{\phi(t-k)}\,dt$ are zero for $|k| > N$. The function $A(\omega)$ with these Fourier coefficients is a *real non-negative trigonometric polynomial of degree N*. Its square root $G(\omega) = \sum g(k)e^{-ik\omega}$, by spectral factorization, is also a polynomial of degree N. The frequency response of the new orthogonalized filter is a ratio of polynomials

$$H_{\text{orth}}(\omega) = \frac{H(\omega)}{G(\omega)}.$$

The input-output equation $y(k) = \sum h_{\text{orth}}(k)x(n-k)$ is an implicit difference equation, from an *autoregressive* moving average filter:

$$\sum_0^N g(k)y(n-k) = \sum_0^N h(k)x(n-k).$$

The new filter is IIR but it only involves $2N+2$ parameters $g(k)$ and $h(k)$. Therefore it can be physically realized, and now the basis has been orthogonalized.

Problem Set 6.4

1. Use the identity $\sin 2\theta = 2\sin\theta\cos\theta$ to show that

$$\left(\cos\frac{\omega}{2}\right)\left(\cos\frac{\omega}{4}\right)\cdots\left(\cos\frac{\omega}{2^N}\right) = \frac{1}{2^N}\,\frac{\sin\omega}{\sin\frac{\omega}{2}}\,\frac{\sin\frac{\omega}{2}}{\sin\frac{\omega}{4}}\cdots\frac{\sin\frac{\omega}{2^{N-1}}}{\sin\frac{\omega}{2^N}}.$$

Cancel sines and let $N\to\infty$ to find a great infinite product:

$$\prod_1^\infty \cos\left(\frac{\omega}{2^j}\right) = \frac{1}{\omega}\sin\omega.$$

2. The Haar filter has $H(\omega) = \frac{1}{2}(1+e^{-i\omega}) = e^{-i\omega/2}\cos\frac{\omega}{2}$. Use $\frac{\omega}{2}$ in Problem 1 to give a new proof for the infinite product (6.47) of $H(\omega/2^j)$:

$$\left(e^{-i\omega/4}\cos\frac{\omega}{4}\right)\left(e^{-i\omega/8}\cos\frac{\omega}{8}\right)\cdots = \left(e^{-i\omega/2}\frac{2}{\omega}\sin\frac{\omega}{2}\right) = \frac{1}{i\omega}(1-e^{-i\omega}).$$

3. Suppose $H(\omega) = \frac{1}{4}(1+e^{-i\omega})^2$. Find $\widehat{\phi}(\omega)$ and $\phi(t)$.

4. If $H(\omega)$ has p zeros at $\omega=\pi$, show that $\widehat{\phi}(\omega)$ has p zeros at $\omega = 2\pi n$ for each $n \neq 0$.

6.5 Biorthogonal Wavelets

This chapter has concentrated on orthogonal wavelets, coming from an orthogonal filter bank. The synthesis filters are transposes of the analysis filters. One multiresolution is all we need. The synthesis wavelets are the same, in this self-orthogonal case, as the analysis wavelets. But from ***biorthogonal filters*** we must expect ***biorthogonal wavelets***.

We now meet a new scaling function $\widetilde{\phi}(t)$. Its translates $\widetilde{\phi}(t-k)$ will span a new lowpass space $\widetilde{V}_0$—different from V_0. There is also a wavelet $\widetilde{w}(t)$. The translates $\widetilde{w}(t-k)$ span a complementary highpass space $\widetilde{W}_0$. The sum of those spaces will be $\widetilde{V}_1 = \widetilde{V}_0 + \widetilde{W}_0$, the next space in the second multiresolution. To a large extent, the theory is achieved by inserting a tilde where appropriate. We want to indicate why this second scale of spaces $\widetilde{V}_j$ is needed.

Biorthogonality comes automatically with inverse matrices. The rows of a 2×2 matrix and the columns of its inverse are biorthogonal:

$$\begin{bmatrix} \text{row} & 1 \\ \text{row} & 2 \end{bmatrix}\begin{bmatrix} \text{column} & \text{column} \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Notice something pleasant. The product in the other order is still I. The right-inverse is also the left-inverse. This order involves *columns times rows*, which are full matrices:

$$\begin{bmatrix} \text{column} \\ 1 \end{bmatrix}\begin{bmatrix} \text{row} & 1 \end{bmatrix} + \begin{bmatrix} \text{column} \\ 2 \end{bmatrix}\begin{bmatrix} \text{row} & 2 \end{bmatrix} = I.$$

Those two column-row products are projections, and they add to I. These simple facts about 2×2 matrices have important parallels for biorthogonal filters and biorthogonal wavelets.

Filter banks display those parallels immediately. The analysis bank has filters $(\downarrow 2)\boldsymbol{H}_0$ and $(\downarrow 2)\boldsymbol{H}_1$. Those are rows 1 and 2. The synthesis bank has expanders before filters, $\boldsymbol{F}_0(\uparrow 2)$ and $\boldsymbol{F}_1(\uparrow 2)$. Those are columns 1 and 2. In the orthogonal case $\boldsymbol{F}_0$ is $\boldsymbol{H}_0^T$ and $\boldsymbol{F}_1$ is $\boldsymbol{H}_1^T$. In the biorthogonal case we don't have transposes but we still have inverses. To understand the pattern for wavelets, we absolutely must return to multiresolution. One scale of spaces $V_0 \subset V_1 \subset \cdots \subset V_j$ is too limited. We need two hierarchies of spaces, V_j in synthesis and $\widetilde{V}_j$ in analysis.

Tilde Notation *Does the tilde go on the analysis functions or the synthesis functions?* Both conventions are equally possible. We hope to agree with other authors! More and more, the tilde is going on the ***analysis*** functions. Then $f(t)$ is expanded in synthesis functions, which have no tilde. But the coefficients come from the analysis functions and have tildes:

$$f_0(t) = \sum \widetilde{a}_{0k}\, \phi(t-k) \text{ is in } V_0, \text{ with } \widetilde{a}_{0k} = \int f(t)\, \widetilde{\phi}(t-k)\, dt \tag{6.52}$$

$$f(t) = \sum\sum \widetilde{b}_{jk}\, w_{jk} \text{ is in } L^2, \text{ with } \widetilde{b}_{jk} = \int f(t)\, \widetilde{w}_{jk}(t)\, dt \tag{6.53}$$

What does this mean for the filter banks that process the coefficients? Those filters use the letter $\boldsymbol{H}$ in analysis and $\boldsymbol{F}$ in synthesis. We will stay with $\boldsymbol{H}$ and $\boldsymbol{F}$ (rather than $\boldsymbol{C}$ and $\boldsymbol{D}$) when discussing biorthogonal filters. An important result in this section is the Fast Wavelet Transform in equation (6.70), and its inverse (the biorthogonal IFWT) in Theorem 6.16.

Biorthogonal Multiresolution

This chapter began with orthogonal bases $\{\phi(t-k)\}$ for V_0 and $\{w(t-k)\}$ for W_0. *The equation* $V_0 \oplus W_0 = V_1$ ***started a multiresolution.*** W_j was the orthogonal (!) complement of V_j inside V_{j+1}. All is well if $\phi(t)$ and $w(t)$ come from an orthogonal bank of FIR filters. Their translates are all orthogonal. They span perpendicular spaces and we have an orthogonal multiresolution.

All is *not* so well if $\phi(t)$ and $w(t)$ fail to have compact support. The filters fail to be FIR. Often this means that we have asked for too much! Instead of orthogonal bases, we should be content with stable bases. An outstanding example is the space of piecewise linear functions. The stable basis consists of the hat function $\phi(t)$ and its translates. That basis is *not orthogonal.*

When the basis is not orthogonal, there is no reason to insist that W_0 must be orthogonal to V_0. If we do, the multiresolution is called ***semi-orthogonal*** in Section 7.4, and we have "prewavelets". But the important property is a stable basis $\{w(t-k)\}$. The highpass coefficients will construct $w(t)$.

Remember the pattern for perfect reconstruction. Coefficients are chosen so that $F_0(z)H_0(z)$ is halfband. When $\phi(t)$ is the hat function from $F_0(z) = \left(\frac{1+z^{-1}}{2}\right)^2$, the other factor $H_0(z)$ needs *five* coefficients. This means $N = 2$ but $\widetilde{N} = 4$. The wavelet has 3-interval support. Then $\phi(t-k)$ and $w(t-k)$ span V_0 and W_0, without orthogonality.

The new *analysis multiresolution* is the point of this section. The coefficients from $H_0(z)$ go into a different dilation equation, whose solution is the ***analyzing function*** $\widetilde{\phi}(t)$:

$$\textbf{Analysis Dilation Equation:} \quad \widetilde{\phi}(t) = \sum_0^{\widetilde{N}} 2h_0(k)\, \widetilde{\phi}(2t-k). \tag{6.54}$$

The coefficients $h_0(k)$ add to 1 as before. The new multiresolution obeys the same conditions as before; just add a tilde. $\widetilde{V}_0$ is spanned by $\{\widetilde{\phi}(t-k)\}$. The space $\widetilde{V}_j$ is spanned by $\{\widetilde{\phi}(2^j t-k)\}$.

They are clearly shift-invariant. The dilation equation (6.54) says that $\widetilde{V}_0 \subset \widetilde{V}_1$. Then also $\widetilde{V}_j \subset \widetilde{V}_{j+1}$. The highpass coefficients produce the wavelet:

$$\textbf{Analysis Wavelet Equation:} \quad \widetilde{w}(t) = \sum_0^N 2h_1(k)\,\widetilde{\phi}(2t-k). \tag{6.55}$$

This wavelet is supported on $[0, \ell]$, where $2\ell = N + \widetilde{N}$. That sum $N + \widetilde{N}$ is the degree of the product filters $F_0(z)H_0(z)$ and $F_1(z)H_1(z)$. Those are symmetric halfband filters, and in the hat function example the degree is $N + \widetilde{N} = 2+4$. Then $\ell = 3$ is odd. The four functions $\phi(t)$, $w(t)$, $\widetilde{\phi}(t)$, $\widetilde{w}(t)$ are graphed in Figure 6.6.

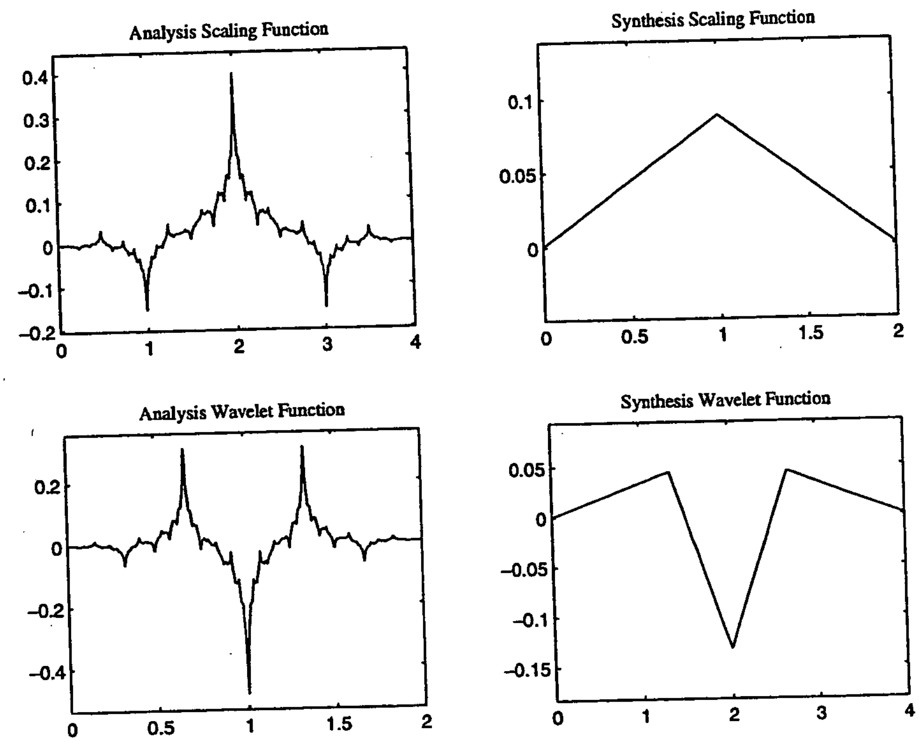


Figure 6.6: Biorthogonal scaling functions and wavelets from a 5/3 PR filter bank.

Biorthogonality in Continuous Time

Our construction of $\phi(t)$, $w(t)$, $\widetilde{\phi}(t)$, $\widetilde{w}(t)$ starts with biorthogonal filters. The lowpass analysi coefficients $h_0(k)$ are *not* double-shift orthogonal to themselves. They are double-shift ***biorthog onal*** to the lowpass synthesis coefficients $f_0(k)$:

$$2\sum h_0(k)f_0(k+2n) = \delta(n). \tag{6.56}$$

This means that $F_0(z)\,H_0(z)$ is halfband. Similarly the highpass filters give $F_1(z)\,H_1(z) =$ half band:

$$2\sum h_1(k)f_1(k+2n) = \delta(n). \tag{6.57}$$

The other key relation is biorthogonality of highpass to lowpass:

$$\sum h_0(k)f_1(k+2n)=0 \quad \text{and} \quad \sum h_1(k)f_0(k+2n)=0. \tag{6.58}$$

The reader knows that all these equations restate perfect reconstruction:

$$F_0(z)=H_1(-z), \quad F_1(z)=-H_0(-z), \quad F_0(z)H_0(z) \textit{ is a halfband filter.}$$

Our question is, *how does biorthogonality appear in continuous time*? The functions $\phi(t)$ and $\widetilde{\phi}(t)$ come from iterating the lowpass filters F_0 and H_0 (with rescaling!). Start the cascade of iterations from $\phi^{(0)}(t)=\widetilde{\phi}^{(0)}(t)=$ box function on $[0,1]$. Their translates have biorthogonality. The box $\phi^{(0)}(t-k)$ has no overlap with $\widetilde{\phi}^{(0)}(t-\ell)$ when $k\neq\ell$. This biorthogonality is preserved at every iteration step, when we use equation (6.56). This is exactly parallel to the earlier proof (6.25) that orthogonality is preserved at each iteration. When $\phi^{(i)}(t)$ and $\widetilde{\phi}^{(i)}(t)$ converge in L^2 to the scaling functions $\phi(t)$ and $\widetilde{\phi}(t)$, those limit functions inherit the same biorthogonality:

$$\int_{-\infty}^{\infty}\phi(t-k)\widetilde{\phi}(t-\ell)\,dt=\delta(k-\ell). \tag{6.59}$$

With $i\to\infty$ in the cascade algorithm, these limits $\phi(t)$ and $\widetilde{\phi}(t)$ solve the synthesis and analysis dilation equations. Now bring in the wavelet equations:

$$\int_{-\infty}^{\infty}\phi(t)\widetilde{w}(t)\,dt=\int_{-\infty}^{\infty}\left(\sum 2h_0(k)\phi(2t-k)\right)\left(\sum 2f_1(\ell)\widetilde{\phi}(2t-\ell)\right)dt. \tag{6.60}$$

That right side is zero because of (6.58) and (6.59). And biorthogonality extends to the translates for the same reason:

$$\int_{-\infty}^{\infty}\phi(t-k)\widetilde{w}(t-\ell)\,dt=0 \quad \text{for all } k \text{ and } \ell. \tag{6.61}$$

Finally the wavelets w and $\widetilde{w}$ are biorthogonal from the wavelet equations and (6.57):

$$\int_{-\infty}^{\infty}w(t-k)\widetilde{w}(t-\ell)\,dt=\delta(k-\ell). \tag{6.62}$$

All this is routine, *provided the cascade algorithms for $\phi(t)$ and $\widetilde{\phi}(t)$ both converge in L^2*. Wavelet theory gives the requirements in Section 7.2, as tests on eigenvalues of two matrices T and $\widetilde{T}$. Suppose those tests are passed (not at all automatic!). The basis functions are biorthogonal. What does this say about the subspaces they span?

Theorem 6.14 *Suppose the filter coefficients satisfy* (6.56)–(6.58) *and also Condition* **E** *(for L^2 convergence of the cascade algorithm). Then the synthesis functions $\phi(t-k)$, $w(t-k)$ are biorthogonal to the analysis functions $\widetilde{\phi}(t-\ell)$, $\widetilde{w}(t-\ell)$ as in* (6.59)–(6.62)*. Each scaling space is orthogonal to the dual wavelet space:*

$$\boxed{V_j \perp \widetilde{W}_j \quad \textit{and} \quad W_j \perp \widetilde{V}_j.} \tag{6.63}$$

V_0 and $\widetilde{W}_0$ are perpendicular because their bases $\phi(t-k)$ and $\widetilde{w}(t-k)$ are perpendicular. When t is replaced by 2^jt, the zero inner products are still zero. (Change variables back to $T=2^jt$.) So at each scale j we have perpendicular subspaces.

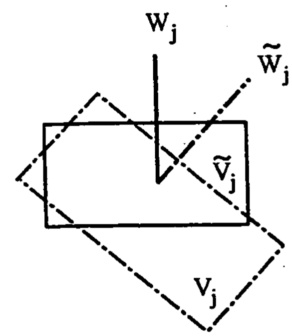


Figure 6.7: V_j is perpendicular to $\widetilde{W}_j$, while $\widetilde{V}_j$ is perpendicular to W_j.

The two multiresolutions are intertwining. As always we have

$$\boxed{V_j + W_j = V_{j+1} \quad \text{and} \quad \widetilde{V}_j + \widetilde{W}_j = \widetilde{V}_{j+1}.} \tag{6.64}$$

These are direct sums but generally not orthogonal sums. The subspaces V_j and W_j have zer intersection, but they are not perpendicular. Instead V_j is perpendicular to $\widetilde{W}_j$. All the subspace $W_{j-1}, W_{j-2}, \ldots$ are then perpendicular to $\widetilde{W}_j$. Similarly all the subspaces $\widetilde{W}_{j-1}, \widetilde{W}_{j-2}, \ldots$ ar perpendicular to W_j. Therefore we have *biorthogonal bases* (*dual bases*):

Corollary *The wavelets $w_{jk}(t) = 2^{j/2}w(2^jt-k)$ and $\widetilde{w}_{jk}(t) = 2^{j/2}\widetilde{w}(2^jt-k)$ are biortho onal bases for L^2:*

$$\int_{-\infty}^{\infty} w_{jk}(t)\widetilde{w}_{JK}(t)\,dt = \delta(j-J)\delta(k-K). \tag{6.6}$$

Representing $f(t)$ in a Wavelet Series

If we have a wavelet basis, we have a wavelet series. Any square-integrable (finite energy) fun tion $f(t)$ can be expanded in wavelets:

$$f(t) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} \widetilde{b}_{jk}\, w_{jk}(t). \tag{6.6}$$

The synthesis wavelets are used to synthesize the function (of course). But the coefficients $\widetilde{b}$ come from inner products with the analysis wavelets. *This is why $\widetilde{b}_{jk}$ has a tilde.* Multiply (6.6 by the analysis wavelet $\widetilde{w}_{JK}(t)$ and integrate over t. Biorthogonality yields

$$\widetilde{b}_{JK} = \int_{-\infty}^{\infty} f(t)\,\widetilde{w}_{JK}(t)\,dt. \tag{6.6}$$

Equations (6.66) and (6.67) are the biorthogonal wavelet transform and its inverse. The tra form takes function to coefficients, the inverse transform synthesizes the function. We show l low how the coefficients can be computed recursively (or pyramidally). This is the *fast* wave transform.

Note that Parseval's equality between $\int |f(t)|^2\,dt$ and $\sum\sum |b_{jk}|^2$ is *not true*. That required orthogonality — and not biorthogonality. When we square and integrate the series for $f(t)$, nonzero inner products come from the products $w_{jk}(t)w_{JK}(t)$. We do have an inequality

$$A\int_{-\infty}^{\infty}|f(t)|^2\,dt \le \sum\sum|\widetilde{b}_{jk}|^2 \le B\int_{-\infty}^{\infty}|f(t)|^2\,dt. \tag{6.68}$$

With $B \ge A > 0$ this says we have a stable basis or ***Riesz basis***. This is true for wavelets, subject to the same Condition **E** [Cohen-Daubechies].

Fast Biorthogonal Wavelet Transform

The reader knows that in practice the wavelet expansion cannot go all the way back to $j = -\infty$. We do not use arbitrarily low frequencies (longer and longer wavelets). A more practical expansion starts with V_0 and $\widetilde{V}_0$, at a scale normalized to $\Delta t = 1$, and goes to V_J and $\widetilde{V}_J$, where the finer scale is 2^{-J}. Enough high-frequency details are included to reproduce an accurate signal.

Thus we work primarily with the subspaces $V_J = V_0 + W_0 + \cdots + W_{J-1}$. There are two important bases for V_J. One is $\phi_{Jk}(t) = 2^{J/2}\phi(2^Jt-k)$ for $-\infty < k < \infty$. These scaling functions are *at level* J. The other basis consists of $\phi_{0k}(t)$ from V_0 and $w_{jk}(t)$ for $0 \le j < J$. Since life is recursive, we are interested first of all in $J = 1$. Then the two ways to expand a signal in V_1 are the scaling basis (fine scale) or scaling functions plus wavelets (coarse scale):

$$\sum \widetilde{a}_{1k}\sqrt{2}\,\phi(2t-k) = \sum \widetilde{a}_{0k}\,\phi(t-k) + \sum \widetilde{b}_{0k}\,w(t-k). \tag{6.69}$$

The key is the pyramid algorithm. This connects $\widetilde{a}_{1k}$ to $\widetilde{a}_{0k}$ and $\widetilde{b}_{0k}$. We are using ϕ and w because this is synthesis of the signal. But the coefficients come from *analysis* of the signal, which uses $\widetilde{\phi}$ and $\widetilde{w}$:

$$\widetilde{a}_{0k} = \int f(t)\,\widetilde{\phi}(t-k)\,dt = \int f(t)\sum_{\ell} h_0(\ell-2k)\,\widetilde{\phi}_{1\ell}(t)\,dt.$$

For $\widetilde{b}_{0k}$ we use the wavelet equation with $h_1(\ell-2k)$ instead of the dilation equation with $h_0(\ell-2k)$. The pyramid has filtering and downsampling:

$$\boxed{\textit{Fast Wavelet Transform}\ \ \widetilde{a}_{0k} = \sum h_0(\ell-2k)\,\widetilde{a}_{1\ell}\ \text{ and }\ \widetilde{b}_{0k} = \sum h_1(\ell-2k)\,\widetilde{a}_{1\ell}.} \tag{6.70}$$

This includes a time-reversal! The same filters H_0^T and H_1^T operate at level j.

That change of basis was not orthogonal. Going backwards, the inverse will not be the transpose. The synthesis filters F_0 and F_1 must do their part.

Theorem 6.15 **Inverse Fast Wavelet Transform:** *The coefficients $\widetilde{a}_{1\ell}$ in the basis $\phi_{1\ell}(t)$ can be computed from $\widetilde{a}_{0k}$ and $\widetilde{b}_{0k}$ by time-reversed synthesis filters:*

$$\boxed{\widetilde{a}_{1\ell} = \textstyle\sum_k f_0(\ell-2k)\,\widetilde{a}_{0k} + \sum_k f_1(\ell-2k)\,\widetilde{b}_{0k}.} \tag{6.71}$$

Proof. Perfect reconstruction operates when (6.70) is substituted into (6.71):

$$\widetilde{a}_{1\ell} = \sum_k f_0(\ell-2k)\sum_m h_0(\ell-2m)\widetilde{a}_{1m} + \sum_k f_1(\ell-2k)\sum_m h_1(\ell-2m)\widetilde{a}_{1m} \tag{6.72}$$

$$\begin{tikzcd}
\tilde{a}_2 \arrow[r, "H_0"] \arrow[dr, "H_1"'] & \tilde{a}_1 \arrow[r, "H_0"] \arrow[dr, "H_1"'] & \tilde{a}_0 \arrow[r, "F_0"] & \tilde{a}_1 \arrow[r, "F_0"] & \tilde{a}_2 \\
 & \tilde{b}_1 & \tilde{b}_0 \arrow[ur, "F_1"'] & \tilde{b}_1 \arrow[ur, "F_1"'] &
\end{tikzcd}$$

Figure 6.8: The pyramid algorithm from $\tilde{a}_2$ back to $\tilde{a}_2$.

The double-shift biorthogonality in (6.56) and (6.57) makes this correct. The beauty of this Mallat algorithm (Figure 6.8 goes up and back down) is the way it connects continuous-time multiresolution to discrete-time filters.

Notice that the whole pyramid operates equally well if $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$ are exchanged with $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$. The dual expansion of $f(t)$ is:

$$f(t) = \sum\sum b_{jk}\, \widetilde{w}_{jk}(t) \quad \text{and} \quad b_{jk} = \int f(t) w_{jk}(t)\, dt. \tag{6.73}$$

All products have a tilde multiplying a non-tilde! This starts with the inverse relation of synthesis to analysis. Tildes can be exchanged with non-tildes throughout (if we want to do it). We will select ϕ and w to be effective in synthesis.

We emphasize one final point, often ignored. The ***coefficients*** in the expansion of $f(t)$ are really different from ***samples*** of $f(t)$. They are inner products, not point values. This distinction is made in Section 7.1.

Filter Construction by Lifting

Herley and Sweldens have proposed (independently) a systematic way to construct biorthogonal filter banks. Only one lowpass filter changes at each step. Starting from short filters, he quickly builds longer ones. In all cases the highpass filters remove aliasing in the standard way: $H_1(z) = F_0(-z)$ and $F_1(z) = -H_0(-z)$. Then equation (4.9) on H_0 and F_0 is the remaining condition for perfect reconstruction. *We drop the subscript zero on these lowpass filters*, and recall the condition (4.9) that removes distortion:

$$F(z)H(z) - F(-z)H(-z) = 2z^{-l} \quad (\text{odd } l). \tag{6.74}$$

Suppose this is satisfied by F and H; the filter bank is PR. *Keeping F fixed, what are the other possible choices for H?* The answer is simple and important:

Theorem 6.16 **(Lifting)** *For fixed $F(z)$, the solutions $H^{\#}(z)$ to (6.74) are*

$$\boxed{H^{\#}(z) = H(z) + F(-z)S(z^2) \quad \textit{for any} \quad S(z).} \tag{6.75}$$

Proof. Substitute $H^{\#}(z)$ to show that equation (6.74) is still satisfied. The new terms are $F(z)F(-z)S(z^2) - F(-z)F(z)S(z^2) = 0$. This is in [HerVet] and [Swel].

Note that with F fixed, *the equation is linear in H*. We are starting from a particular solution (right side $= 2z^{-l}$). To this particular H we are adding solutions $F(-z)S(z^2)$ to the homogeneous equation (right side $=$ zero). Thus the even $S(z^2)$ displays the degrees of freedom that remain when the PR condition is satisfied (Problem 5). That freedom is used in the Daubechies construction to achieve zeros at $z = -1$.

We still want and need zeros at -1, which is $\omega = \pi$. We also need stable bases. (Section 7.2 will state the stability requirement as Condition E. There is not yet a simple way to decide which $S(z^2)$ are permitted by this condition. In practice, we construct a potentially useful $H^{\#}(z)$ and test it for Condition E.) This section will build in other important properties — *linear phase, interpolating scaling functions, binary* (dyadic) *filter coefficients*. Those come at the expense of higher-order zeros at π.

Dual lifting is also useful. In this case we fix the analysis filter $H(z)$. The PR condition (6.74) becomes linear in $F(z)$. The lifted solutions are

$$\boxed{F^{\#}(z) = F(z) + H(-z)T(z^2) \quad \text{for any} \quad T(z).} \tag{6.76}$$

Starting from the Haar filter or even from the "Lazy filter", we alternate lifting and dual lifting to construct high-order biorthogonal filter banks with good properties. All these filters would be attainable directly from the (second-degree) PR equations. Lifting is a way to solve them as a sequence of linear equations, with F or H fixed at each step. Then we control more closely the final result.

Sweldens also emphasizes that lifting yields a *faster implementation* of the wavelet transform and its inverse. The Mallat filter tree, which is subband filtering, is climbed in smaller steps. This is related to a lattice factorization.

Example 6.6. The Lazy filter has $H(z) = 1$ and $F(z) = z^{-1}$. There is no true filtering, only subsampling from $(\downarrow 2)$. Section 4.3 displayed block diagrams of this filter bank. Its polyphase matrix is $\boldsymbol{H}_p = \boldsymbol{I}$.

Suppose we keep $H(z) = 1$ and apply dual lifting to the synthesis filter:

$$F^{\#}(z) = z^{-1} + T(z^2). \tag{6.77}$$

$F^{\#}(z)$ can be any halfband filter, with one odd power $z^{-l} = z^{-1}$. We are allowing S and T to contain powers of z as well as z^{-1}. This is needed in order to create symmetric filters.

Earlier we centered the product $H(z)F(z)$, multiplying by z^l. In this case centering gives $zF^{\#}(z) = 1 + zT(z^2)$. Then D_4 comes from $T(z) = (-z + 9 + 9z^{-1} - z^2)/16$. Every maxflat halfband filter D_{2p} can be lifted and centered from the Lazy filter. This section will create symmetric biorthogonal filters, by lifting $H = 1$ while $F = D_{2p}$.

Note the scaling functions for this important example. In analysis we have the delta function $\widetilde{\phi}(t) = \delta(t)$ from $H(z) = 1$. This solves the dilation equation $\delta(t) = 2\delta(2t)$. In synthesis $F^{\#}$ yields an ***interpolating scaling function***, with $\phi(n) = \delta(n)$. This is certainly biorthogonal to the analysis functions!

$$\int \phi(t)\widetilde{\phi}(t-n)\,dt = \int \phi(t)\delta(t-n)\,dt = \phi(n) = \delta(n). \tag{6.78}$$

You can see in another way that $\phi(n) = \delta(n)$, because the values of ϕ at the integers come from the $\lambda = 1$ eigenvector of $(\downarrow 2)2F^{\#}$. When the filter is halfband, the center column of the matrix $2F^{\#}$ is the vector δ. This is an eigenvector with $\lambda = 1$. Assuming a stable basis, there are no other eigenvectors for $\lambda = 1$ by Condition E. So $\phi(n)$ agrees with $\delta(n)$.

This interpolating property is highly useful in several applications. But the analysis filter with $H(z) = 1$ and $\widetilde{\phi}(t) = \delta(t)$ is generally not acceptable. *Therefore we now lift H. That produces a new pair $(H^{\#}, F^{\#})$ which seems extremely promising.*

Biorthogonal Filters with Binary Coefficients

A *binary coefficient* or *dyadic coefficient* is an integer divided by a power of 2. The maxflat halfband filters $D_{2p}(z)$ all have binary coefficients. This is clear from the Daubechies formula (5.75), where the binomial coefficients are integers. Multiplication by a binary number can be executed entirely by *shifts* and *adds*. Roundoff error is eliminated. And on some architectures, the filter needs less time and less space.

We are therefore highly interested in binary filters. Most factorizations of D_{2p} — this has been our route to orthogonal and biorthogonal filters — are *not* binary. There are zeros at irrational points like $z = 2-\sqrt{3}$. But we can certainly move zeros at $z=-1$ between analysis and synthesis. This operation we call ***balancing***.

Moving $(\frac{1+z^{-1}}{2})$ *from* $F(z)$ *to* $H(z)$ *maintains binary coefficients and symmetry*:

$$\boldsymbol{h}_{new}(n) = \tfrac{1}{2}[\boldsymbol{h}_{old}(n) + \boldsymbol{h}_{old}(n-1)] \quad \text{and} \quad \boldsymbol{f}_{new}(n) = \tfrac{1}{2}[\boldsymbol{f}_{old}(n) - \boldsymbol{f}_{new}(n-1)]. \tag{6.79}$$

Note $\boldsymbol{f}_{new}$ at the end. We are dividing $F(z)$ by $(\frac{1+z^{-1}}{2})$. The product $F_{new}H_{new}$ equals $F_{old}H_{old}$, so biorthogonality is preserved. The scaling function $\widetilde{\phi}_{new}(t)$ is $\widetilde{\phi}_{old}(t)$ convolved with the box function. Therefore it has exactly one more derivative than $\widetilde{\phi}_{old}(t)$ (Section 7.3). Similarly $\phi_{new}(t)$ from $\boldsymbol{F}_{new}$ has one less derivative than $\phi_{old}(t)$ from F_{old}. In a filter bank we avoid the destructive factors $\sqrt{2}$, by putting both of them into synthesis. Our convention below is $H(1) = 1$ and $F(1) = 2$.

Our example $\boldsymbol{h1} = [1]$ and $\boldsymbol{f7} = [-1 \;\; 0 \;\; 9 \;\; 16 \;\; 9 \;\; 0 \;\; -1]/16$ is binary. This 1/7 filter bank has denominator 16. *Balancing will produce* 2/6 *and* 3/5, *still binary and symmetric*:

$$\boldsymbol{h2} = [1 \;\; 1]/2 \text{ and } \boldsymbol{f6} = [-1 \;\; 1 \;\; 8 \;\; 8 \;\; 1 \;\; -1]/8 \quad \text{with 1/3 zeros at } \pi$$

$$\boldsymbol{h3} = [1 \;\; 2 \;\; 1]/4 \text{ and } \boldsymbol{f5} = [-1 \;\; 2 \;\; 6 \;\; 2 \;\; -1]/4 \quad \text{with 2/2 zeros at } \pi.$$

These are very effective short filters. They are probably the best — after reversing the secon pair to 5/3. In the experiments of Chapters 10 and 11 they are comparable and quite effective As factors of the maxflat D_6 filter, we have seen them before. An interesting feature of $\boldsymbol{f5}$ i that its scaling function $\phi(t)$ is infinite at all binary points! Section 7.3 confirms that this $\phi(t$ nevertheless has finite energy (and even 0.44 derivatives in the energy sense). By removing tw zeros at $z=-1$ from $\boldsymbol{f7}$, we have stolen its smoothness. Enough is left to make it good amon short filters (but moved to the analysis side).

For serious compression we need more zeros — which means longer filters. Our previou route was to factor a long maxflat filter. When one factor is $F(z) = (\frac{1+z^{-1}}{2})^p$ the pair is sti binary and symmetric. The scaling function $\phi(t)$ for this factor is a ***spline***. It has maximu smoothness for its length (p intervals) coming from a maximum number of zeros at π (p zeros These are outstanding filters, when we keep enough smoothness in the other factor. Taking o three zeros from D_6 would leave a filter $[-1\;3\;3\;-1]/2$ which has one zero but negative smoot ness — too risky to iterate. Taking three zeros from D_8 is allowed. Now, instead of factoring long filter, we will lift a short filter.

Lifting will not maximize the number of zeros at π (although we like those zeros). Our fi lifting will go from 1/7 to 9/7, and we choose $S(z^2)$ in (6.74) to give two zeros at π. Here is t result of lifting $H=1$ (all filters are symmetric):

$$\boldsymbol{h9} = [1 \;\; 0 \;\; -8 \;\; 16 \;\; 46 \;\ldots]/64 \text{ and } \boldsymbol{f7} = [-1 \;\; 0 \;\; 9 \;\; 16 \;\cdots]/16 \;\; (2/4 \text{ zeros}).$$

These are binary filters. Balancing the zeros by (6.79) yields 10/6 from 9/7:

$$\boldsymbol{h10} = [1 \;\; 1 \;\; -8 \;\; 8 \;\; 62 \;\cdots]/128 \text{ and } \boldsymbol{f6} = [-1 \;\; 1 \;\; 8 \;\; 8 \;\; 1 \;\; -1]/8 \text{ (3/3 zeros).}$$

This 10/6 pair gives better compression as 6/10—reversing analysis and synthesis. So does 5/11 from 11/5, after another balancing (or unbalancing!) step:

$$\boldsymbol{h11} = [1 \;\; 2 \;\; -7 \;\; 0 \;\; 70 \;\; 124 \;\cdots]/256 \text{ and } \boldsymbol{f5} \text{ above (4/2 zeros).}$$

Figure 6.9 shows scaling functions $\widetilde{\phi}(t)$ and $\phi(t)$ for analysis and synthesis. You can see how the zeros affect the smoothness. You cannot easily see which pair is best in compression—that depends on the image.

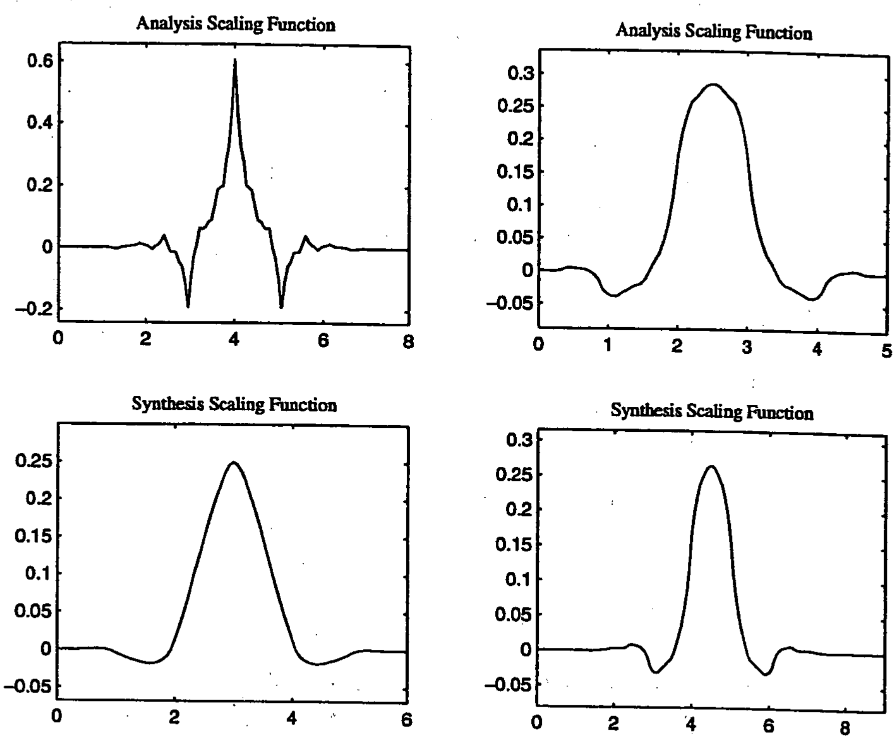


Figure 6.9: Scaling functions for $\boldsymbol{h9/f7}$ and $\boldsymbol{h6/f10}$.

Note 1 The reader might be interested in the construction of these new (1995) binary filters. The first author found the 9/7 pair in September, but not by lifting. With $\boldsymbol{f7}$ fixed, he solved the halfband equation $(\downarrow 2)(\boldsymbol{f7} * \boldsymbol{h9}) = \boldsymbol{\delta}$ for the symmetric filter $\boldsymbol{h9}$ with two zeros at π. (The first zero determines the middle coefficient from the others; the second zero is automatic by symmetry.) When reporting this result for the *Wavelet Digest*, he learned that Wim Sweldens had created a whole family of *binary symmetric filters* earlier in 1995 by lifting. We propose to call them "binlets". Here are the next filters $\boldsymbol{h13/f7}$ and $\boldsymbol{13h/f11}$. All signs indicate that $\boldsymbol{h13/f7}$ is the right choice:

$$\boldsymbol{h13} = [-1 \;\; 0 \;\; 18 \;\; -16 \;\; -63 \;\; 144 \;\; 348 \;\cdots]/512 \text{ with } \boldsymbol{f7} \text{ (4/4 zeros)}$$

$$13h = [-3\ 0\ 22\ 0\ -125\ 256\ 724\ 256\ -125\ 0\ 22\ 0\ -3]/1024 \quad \text{with}$$

$$f11 = D_6 = [3\ 0\ -25\ 0\ 150\ 256\ 150\ 0\ -25\ 0\ 3]/256 \quad (2/6 \text{ zeros}).$$

Extra length gives more zeros and higher compression, up to a point. *Then ringing destroys the image quality*. See Sections 10.1 and 11.2 for the artifacts that plague image compression. The boats in Figure 7.4 offer a visual comparison.

Note 2 [Majani2] emphasizes the importance of a *reversible integer implementation*. Integer inputs are reconstructed exactly. Orthogonal transforms seem not to be reversible (except Haar for $M = 2$ channels and Hadamard for certain $M > 2$). The 2/6 biorthogonal transform with $\boldsymbol{h2} = [1\ \ 1]$ is reversible and very useful. Lowpass components y_{2i} come first:

$$y_{2i} = x_{2i} + x_{2i+1} \quad \text{and then} \quad y_{2i+1} = x_{2i+1} - \lfloor y_{2i}/2 \rfloor + \lfloor (y_{2i-2} - y_{2i+2})/16 \rfloor.$$

The inverse also has "even = even + f(odd)" and "odd = odd + g(even)":

$$x_{2i} = y_{2i} - x_{2i+1} \quad \text{and then} \quad x_{2i+1} = y_{2i+1} + \lfloor y_{2i}/2 \rfloor - \lfloor (y_{2i-2} - y_{2i+2})/16 \rfloor.$$

Majani has shown that the new binary 9/7 and 13/7 transforms have reversible forms (lossless in integers). The normalized DCT is not reversible for integer data.

Note 3 The maxflat Daubechies filters with $4p-1$ coefficients and $2p$ zeros are also known as Deslauriers-Dubuc filters [DesDub, CDM]. They interpolate because they are halfband. They leave the values $x(n)$ unchanged and produce midpoint values $x(n+\frac{1}{2})$. Section 5.5 confirmed that all polynomials of degree less than $2p$ are interpolated exactly. Recursive subdivision starting from $x(n) = \delta(n)$ converges to the scaling function $\phi(t)$ by the cascade algorithm!

Problem Set 6.5

1. Double-shift orthogonality of lowpass filters is $2\sum h_0(k) f_0(k+2n) = \delta(n)$. Show that in frequency this becomes

$$H_0(\omega)F_0(\omega) + H_0(\omega+\pi)F_0(\omega+\pi) = 1.$$

Write the same equation in the z-domain.

2. Problem 1 involves a row and column of the modulation matrices $\boldsymbol{F}_m$ and $\boldsymbol{H}_m$:

$$\boldsymbol{F}_m(z)\boldsymbol{H}_m(z) = \begin{bmatrix} F_0(z) & F_1(z) \\ F_0(-z) & F_1(-z) \end{bmatrix} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix}.$$

Which row-column multiplications correspond to which equations (6.56)–(6.58)?

3. Suppose $H_0(\omega)F_0(\omega) + H_0(\omega+\pi)F_0(\omega+\pi) = 1$ as required. By alternating flip

$$H_1(\omega) = e^{-i\omega}F_0(\omega+\pi) \quad \text{and} \quad F_1(\omega) = -e^{-i\omega}H_0(\omega+\pi).$$

Show that the other entries of $\boldsymbol{F}_m(z)\boldsymbol{H}_m(z) = \boldsymbol{I}$ are then correct.

4. What wavelets come from the biorthogonal filters with $H_0 = 1$, $F_0 = \frac{1}{2}z + 1 + \frac{1}{2}z^{-1}$, $H_1 = \frac{1}{2}z - 1 + \frac{1}{2}z^{-1}$, $F_1 = -1$? Recognize the delta and hat:

$$\widetilde{\phi}(t) = 2\widetilde{\phi}(2t) \quad \text{and} \quad \phi(t) = \tfrac{1}{2}\phi(2t+1) + \phi(2t) + \tfrac{1}{2}\phi(2t-1).$$

Then construct wavelets from $\widetilde{w}(t) = -\frac{1}{2}\widetilde{\phi}(2t+1) + \widetilde{\phi}(2t) - \frac{1}{2}\widetilde{\phi}(2t-1)$ and $w(t) = 2\phi(2t-1)$. Check the biorthogonality conditions

$$\begin{aligned}
\int \phi(t)\widetilde{\phi}(t-k)\,dt &= \int w(t)\widetilde{w}(t-k)\,dt = \delta(k) \quad \text{and} \\
\int \phi(t)\widetilde{w}(t-k)\,dt &= \int \widetilde{\phi}(t)w(t-k)\,dt = 0.
\end{aligned}$$

5. Lifting from the Haar filters $H(z) = \frac{1}{2}(1+z^{-1})$ and $F(z) = 1+z^{-1}$, show that all PR solutions have the form (6.75). The difference $D = H^{\#} - H$ must satisfy $D(z)F(z) = D(-z)F(-z)$ from (6.74). Substitute $D(z) = a + bz^{-1} + cz^{-2} + dz^{-3} + \cdots$ to show that it has the form $F(-z)S(z^2)$.

6. Lifting (H, F) to $(H^{\#}, F)$ does not change the scaling function $\phi(t)$. Show that the new wavelet is $w^{\#}(t) = w(t) - \sum s(k)\phi(t-k)$.

7. The *fast wavelet transform* is subband filtering of the inner products $a_{jk} = \langle f(t), \widetilde{\phi}_{jk}(t)\rangle$. The highpass channel produces $b_{jk} = \langle f(t), \widetilde{w}_{jk}(t)\rangle$ by (6.70):

$$a_{jk} = \sum h(l-2k)a_{j+1,l} \quad \text{and} \quad b_{jk} = \sum h_1(l-2k)a_{j+1,l}.$$

Suppose H is lifted to $H^{\#}$. *Show that the lifted* $a^{\#}_{jk}$ *are* $a^{\#}_{jk} = a_{jk} + \sum s(l-k)b_{jl}$. The fast *inverse* transform unlifts by $-s(l-k)$ in the same way. Then it inverts the (H, H_1) transform as usual by $a_{j+1,k} = \sum f(k-2l)a_{jl} + \sum f_1(k-2l)b_{jl}$.

8. From the input $x(n) = a_{1k}$ compute even samples $a_{0k} = a_{1,2k}$ and odd $b_{0k} = a_{1,2k+1} - (a_{0k} + a_{0,k+1})$. Then lift to $a^{\#}_{0k} = a_{0k} + (b_{0,k-1} + b_{0k})/4$. Combine to recognize the 5/3 filter bank, computed more efficiently and in place—no auxiliary memory.

9. Which filter h gives linear interpolation at each step of recursive subdivision? What is $\phi(t)$?

Appendix 2

Wavelets and Dilation Equations

Gilbert Strang

Siam Review 31 (1989) 613–627

Abstract

Wavelets are new families of basis functions that yield the representation $f(x) = \sum b_{ij} W(2^j x - k)$. Their construction begins with the solution $\phi(x)$ to a dilation equation with coefficients c_k. Then W comes from ϕ, and the basis comes by translation and dilation of W. It is shown in Part 1 how conditions on the c_k lead to approximation properties and orthogonality properties of the wavelets. Part 2 describes the recursive algorithms (also based on the c_k) that decompose and reconstruct f. The object of wavelets is to localize as far as possible in both time and frequency, with efficient algorithms.

Wavelets are based on translation ($W(x) \to W(x+1)$) and above all on dilation ($W(x) \to W(2x)$). It is remarkable how long it has taken for "*dilation equations*" to be mentioned beside differential equations and difference equations. True, they are hardly in the same league. But ideas about wavelets are coming fast. The mathematics is attractive and several important applications seem to fit—I hope this survey will be helpful. You should know that its author is neither an expert nor an evangelist.

The goal is a new way to represent functions—especially functions that are local in time and frequency (or space and wave number). Compare with Fourier series. Sines and cosines are perfectly local in frequency, but global in x or t. A short pulse has slowly decaying coefficients that are hard to measure. To reconstruct the pulse, a Fourier series depends heavily on cancellation. The whole of Fourier analysis, relating properties of functions to properties of coefficients, is made difficult (some say interesting) by the nonlocal support of $\sin x$.

In achieving local support we lose the greatest property of the basis $\{e^{inx}\}$. With respect to a wavelet basis the differentiation operator is not diagonal. Wavelets are not eigenfunctions of $\partial/\partial x$, and frequencies are mixed up. The uncertainty principle imposes limits on what is possible in x and ξ together. The commutator $(\partial/\partial x)(\partial/\partial\xi) - (\partial/\partial\xi)(\partial/\partial x)$ is a multiple of the identity (since $(\partial/\partial x)(xu) - x(\partial u/\partial x) = u$), so we cannot diagonalize both operators. But a good "microlocalization" leaves $\partial/\partial x$ nearly diagonal, and at the same time nearly diagonalizes $\partial/\partial\xi$ (which is multiplication by x). To connect dilation with multiplication by x, differentiate $f(cx)$ with respect to c at $c = 1$.

The second important property of $\{e^{inx}\}$ is orthogonality. That can be saved. Wavelets can be made orthogonal to their own dilations and translations. Then $\int W(x)W(2^j x - k)\,dx = 0$ for all integers j and k. The wavelet basis has two indices, in which k is translation and j is dilation

or compression. It suggests multigrid. A wavelet expansion $\sum b_{jk} W_{jk}(x)$ is a *multiresolution* of $f(x)$, in which b_{jk} carries information about f near $\xi = 2^j$ and $x = 2^{-j}k$. The sum on k is the detail at the scaling level $h = 2^{-j}$.

Orthogonality is not easy to achieve with local support. Truncated at zero and 2π, a sine wave $\phi(x)$ is orthogonal to $\phi(2x)$ but not to $\phi(4x)$. The "windowed Fourier transform" combines smoothness with local support by bringing $e^{i\xi x}$ gradually to zero, but it is not fully satisfactory. The price of orthogonality with compact support is irregular basis functions. We live with these wavelets by doing all computations recursively (this subject is recursion heaven). And it is important to recognize that orthogonality and even linear independence (!) are not essential in the representation of functions. Wavelets need not be orthogonal.

This brief introduction cannot do justice to the applications. Nor can we attempt a proper history—it would be mostly in French. The idea of wavelets grew out of seismic analysis. Their development has been led by Yves Meyer, whose book will describe a new chapter in harmonic analysis (connecting to work of Calderòn, Grossmann, Morlet, Coifman, Weiss, and many others). The interest in wavelets is both pure and applied—like the interest in splines.

Part 1 of this paper establishes the properties of wavelets—approximation through Condition A and orthogonality through Condition O. Since we never see wavelets as functions (only recursively), their properties have to be discovered indirectly. We absolutely need these properties in order to have any idea what the algorithms are producing. Then Part 2 begins with a piecewise constant example (ϕ is a box function, the wavelet is Haar's). The example reveals a lot with no deep analysis. You could go directly to Part 2, about algorithms, and then return to dilation equations.

1. Dilation Equations: Construction of ϕ

The basic dilation equation is a two-scale difference equation:

$$\phi(x) = \sum c_k \phi(2x - k). \tag{A.1}$$

We look for a solution normalized by $\sum \phi\, dx = 1$. The first requirement on the coefficients c_k comes from multiplying by 2 and integrating:

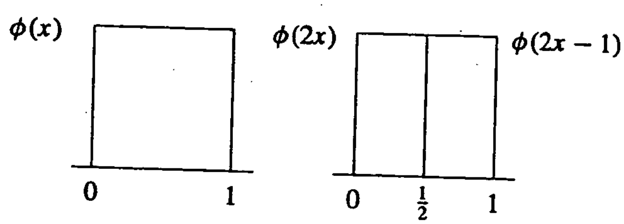
$$2 \int \phi\, dx = \sum c_k \int \phi(2x - k)\, d(2x - k) \quad \text{yields} \quad \sum c_k = 2.$$

Uniqueness of ϕ is ensured by $\sum c_k = 2$. A smooth solution is not ensured. For a striking example, set $c_0 = 2$:

$$\text{The delta function} \quad \phi = \delta \quad \text{satisfies} \quad \delta(x) = 2\delta(2x).$$

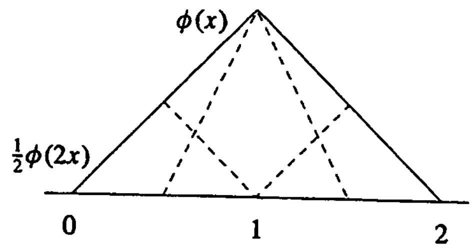
That dilation of δ is unfamiliar (but somehow very pleasing). For other c's, spline functions appear:

Box function:



$c_0 = 1$
$c_1 = 1$

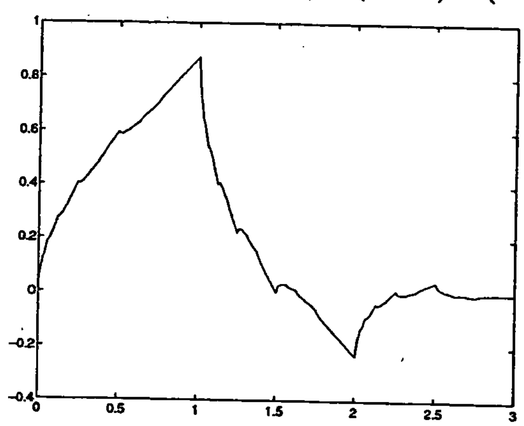
Hat function:



$c_0 = \frac{1}{2}$
$c_1 = 1$
$c_2 = \frac{1}{2}$

We now outline three constructions of the "scaling function" ϕ. Those constructions display very clearly the mathematics of dilation. Then we turn to wavelets, their properties and their purpose. A wavelet $W(x)$ is a second combination (involving the same recursion coefficients c_k) of the translates $\phi(2x - k)$.

Construction 1. Iterate $\phi_j(x) = \sum c_k \phi_{j-1}(2x - k)$ with the box function as $\phi_0(x)$. When $c_0 = 2$ the boxes get taller and thinner, approximating the delta function. For $c_0 = c_1 = 1$ the box is invariant: $\phi_j = \phi_0$. For $\frac{1}{2}, 1, \frac{1}{2}$ the hat function appears as $j \to \infty$, and $\frac{1}{8}, \frac{4}{8}, \frac{6}{8}, \frac{4}{8}, \frac{1}{8}$ yields the cubic B-spline. An example that will be important (an inspiration of Daubechies — we propose the notation D_4) has coefficients $\frac{1}{4}\left(1+\sqrt{3}\right), \frac{1}{4}\left(3+\sqrt{3}\right), \frac{1}{4}\left(3-\sqrt{3}\right)$, and $\frac{1}{4}\left(1-\sqrt{3}\right)$:



This scaling function D_4 leads to orthogonal wavelets. *It is not as smooth as it looks.* Note that the Weierstrass nowhere differentiable function, which is $\sum b^n \cos(3^n x)$, involves dilation by 3. So does de Rham's function, which has $c_k = \frac{2}{3}, \frac{1}{3}, 1, \frac{1}{3}, \frac{2}{3}$ adding to 3. Resnikoff has found a connection between Weierstrass functions and wavelets.

Construction 2. The second construction takes the Fourier transform of (1):

$$\begin{aligned} \widehat{\phi}(\xi) &= \sum c_k \int \phi(2x - k) e^{i\xi x}\, dx \\ &= \tfrac{1}{2}\left(\sum c_k e^{ik\xi/2}\right) \int \phi(y) e^{iy\xi/2}\, dy = P\left(\frac{\xi}{2}\right)\widehat{\phi}\left(\frac{\xi}{2}\right). \end{aligned} \tag{A.2}$$

The symbol $P(\xi) = \frac{1}{2}\sum c_k e^{ik\xi}$ is the crucial function in this theory. Note that $P(0) = 1$. Now repeat (2) at $\xi/2, \xi/4, \ldots$ and recall $\widehat{\phi}(0) = \int \phi\, dx = 1$:

$$\widehat{\phi}(\xi) = \left[\prod_{1}^{n} P\left(\frac{\xi}{2^j}\right)\right]\widehat{\phi}\left(\frac{\xi}{2^N}\right) \quad \text{approaches} \quad \prod_{1}^{\infty} P\left(\frac{\xi}{2^j}\right). \tag{A.3}$$

For $c_0 = 2$ we find $P \equiv 1$ and $\widehat{\phi} \equiv 1$, the transform of the delta function. For $c_0 = c_1 = 1$ the products of the P's are geometric series:

$$P\left(\frac{\xi}{2}\right)P\left(\frac{\xi}{4}\right) = \tfrac{1}{4}\left(1 + e^{i\xi/2}\right)\left(1 + e^{i\xi/4}\right) = \frac{1 - e^{i\xi}}{4\left(1 - e^{i\xi/4}\right)}.$$

As $N \to \infty$ this approaches the infinite product $\left(1 - e^{i\xi}\right)/(-i\xi)$. This is $\int_0^1 e^{i\xi x}\,dx$, the transform of the box function. The hat function comes from squaring $P(\xi)$ which by (3) also squares $\widehat{\phi}(\xi)$. (Multiplication of P's is $\frac{1}{2}$ times convolution of c's). The cubic B-spline comes from squaring again.

Construction 3. This construction of ϕ works directly with the recursion. Suppose ϕ is known at the integers $x = j$. The recursion (1) gives ϕ at the half-integers. Then it gives ϕ at the quarter-integers, and ultimately at all dyadic points $x = k/2^j$. This is fast to program. *All good wavelet calculations use recursion.*

The values of ϕ at the integers come from an eigenvector. With the four Daubechies coefficients, set $x = 1$ and $x = 2$ in the dilation equation (1) and use the fact that $\phi = 0$ unless $0 < x < 3$:

$$\begin{aligned} \phi(1) &= \tfrac{1}{4}(3+\sqrt{3})\phi(1) + \tfrac{1}{4}(1+\sqrt{3})\phi(2) \\ \phi(2) &= \tfrac{1}{4}(1-\sqrt{3})\phi(1) + \tfrac{1}{4}(3-\sqrt{3})\phi(2). \end{aligned} \tag{A.4}$$

This is $\phi = L\phi$, with matrix entries $L_{ij} = c_{2i-j}$. Compare with c_{i-j} for an ordinary difference equation. The eigenvalues are 1 and $\frac{1}{2}$. The eigenvector for $\lambda = 1$ has components $\phi(1) = \frac{1}{2}\left(1+\sqrt{3}\right)$ and $\phi(2) = \frac{1}{2}\left(1-\sqrt{3}\right)$, which are the heights on our graph of D_4. The other eigenvalue $\lambda = \frac{1}{2}$ means that the recursion can be differentiated: $\phi'(x) = \sum c_k 2\phi'(2x-k)$ leads similarly to $\phi'(1)$ and $\phi'(2)$. In some weak sense, $\phi = D_4$ has a "dilational derivative." For the hat function, the recursion matrix (see below) again has $\lambda = 1, \frac{1}{2}$. For the cubic spline the eigenvalues are $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$.

To repeat for emphasis: From $\phi(1)$ and $\phi(2)$ the recursion gives everything.

In these constructions the properties of $P(\xi) = \frac{1}{2}\sum c_k e^{ik\xi}$ are decisive. The precise hypotheses are in flux, and infinitely many c_k can be allowed. One basic property will bring together the theory of dilation equations, before we go on to wavelets.

1.1. Dilation equations: Fundamental theorem. The accuracy of piecewise polynomial approximation, by splines or finite elements, depends on the answer to this question: To what degree $p-1$ can the polynomials $1, x, x^2, \ldots, x^{p-1}$ be reproduced exactly by the approximating functions? When the polynomials are "in the space," the approximation error is of order h^p. In our case, the approximating functions are $\phi(x)$ and its translates. Splines are the best at approximation, and finite elements have the narrowest support — but both are weeded out when we require orthogonality. There is already a theory of approximation by translates. It connects p with the properties of $\widehat{\phi}$. The link is the Poisson summation formula. When ϕ solves a dilation equation, that throws new questions into the theory — it is extremely satisfying that these new questions have the same answers.

For approximation with accuracy h^p, the Fourier transform $\widehat{\phi}$ must have zeros of order p at all points $\xi = 2\pi n$ (except at $\xi = 0$ where $\widehat{\phi} = 1$). Notice how easily that converts to a condition on the symbol P. According to (3), the transform $\widehat{\phi}$ is the infinite product of $P\left(\xi/2^j\right)$. At $\xi = 2\pi$ the first factor is $P(\pi)$. At $\xi = 4\pi$ the second factor becomes $P(\pi)$. At $\xi = 6\pi$ the first factor is $P(3\pi)$, which by periodicity is the same as $P(\pi)$. The zeros of P produce zeros of $\widehat{\phi}$:

Condition A. The symbol $P = \frac{1}{2}\sum c_k e^{ik\xi}$ has a zero of order p at $\xi = \pi$. Equivalently, the coefficients c_k satisfy the sum rules that yield $P^{(m)}(\pi) = 0$:

$$\sum(-1)^k k^m c_k = 0, \quad m = 0, 1, \ldots, p-1. \tag{A.5}$$

The box function has $P = \frac{1}{2}\left(1 + e^{i\xi}\right)$ and $p = 1$. The hat function has $p = 2$ and so does D_4. The cubic spline has $p = 4$.

A zero at $\xi = \pi/2$ (instead of π) would also produce the desired zeros in the product $\widehat{\phi}$. Thus Condition A is not strictly necessary in what follows. Choosing $c_0 = 1$ and $c_2 = 1$ and $P = \frac{1}{2}\left(1 + e^{2i\xi}\right)$ stretches out the box function—it becomes $\phi = \frac{1}{2}$ on the double interval $0 < x \leq 2$. But $P(\pi/2) = 0$ produces instability and linear dependence—the alternating sum of stretched boxes is $\sum(-1)^k\phi(x-k) = 0$. With the added requirement of stability, the condition is exactly right.

The fundamental theorem states the consequences of Condition A:

1. The polynomials $1, x, \ldots, x^{p-1}$ are linear combinations of the translates $\phi(x-k)$.

2. Smooth functions can be approximated with error $O(h^p)$ by combinations at every scale $h = 2^{-j}$:
$$\left\| f - \sum_k a_k \phi(2^j x - k) \right\| \leq C 2^{-jp} \left\| f^{(p)} \right\| \quad \text{for suitable } a_k.$$

3. The first p moments of the wavelet $W(x)$ (see below) are zero:
$$\int x^m W(x)\,dx = 0 \quad \text{for } m = 0, \ldots, p-1.$$

4. The wavelet coefficients of a smooth function decay like $\left|\int f(x) W\left(2^j x\right) dx\right| \leq C 2^{-jp}$.

5. The recursion matrix M_N determining ϕ at the integers has eigenvalues $1, \frac{1}{2}, \ldots, \left(\frac{1}{2}\right)^{p-1}$.

1 and **2** come from approximation theory. The combination of ϕ's at scale j is also a combination $\sum b_{jk} W\left(2^j x - k\right)$ down to scale j. **3** and **4** are easy once wavelets are defined. Mallat gives a sharp result, with properly stated requirements on the smoothness and decay of ϕ: The H^p norm of f is equivalent to the corresponding norm of its coefficients b_{jk}. Wavelets lead to unconditional bases, suitable for a wide range of function spaces.

It is **5** that makes $\phi(x)$ smoother as p increases and also makes the constructions successful. The smoothness is weaker than $\phi \in C^{p-1}$, but it is striking that "dilational derivatives" come at the same time as higher degrees of approximation. What remains to be studied is orthogonality—which imposes an entirely different condition on the c_k.

Remark 1. Suppose the basic recursion has coefficients $c_0, \ldots, c_N$. Then ϕ is zero outside the interval $[0, N]$. With continuity it follows that $\phi(0) = 0$ and $\phi(N) = 0$. Those were assumed in (4) when we determined $\phi = D_4$ at the integers. For the box function with $N = 1$, $\phi(0)$ and $\phi(N)$ cannot both be dropped. Our recursion matrix will be $(M_N)_{ij} = c_{2i-j}$ with $i, j = 0, \ldots, N-1$. For the box function $M_1 = [1]$ has eigenvalue $\lambda = 1$, as expected in **5** above.

The spectrum of the infinite matrix M (allowing all i, j) is an attractive problem in operator theory. Notice that M is *convolution followed by decimation*—multiplication by the matrix c_{i-j} followed by projection onto even-numbered coordinates. By contrast with the usual Toeplitz case, eigenfunctions can have compact support! Homogeneous difference equations with zero boundary conditions lead to $\phi = 0$, but not so for dilation equations.

***Remark* 2.** The minimum requirement is $p = 1$. Then $P(\pi) = 0$, which means that $\sum c_{2k} = \sum c_{2k+1}$. Since $\sum c_k = 2$, the columns of M add to 1:

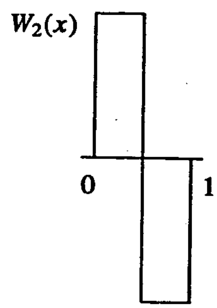
$$M_N = \begin{bmatrix} c_0 & & & \\ c_2 & c_1 & c_0 & \\ & c_3 & c_2 & c_1 \\ & & & c_3 \end{bmatrix} \qquad \begin{array}{l}\text{steps of 2 down columns}\\ \text{steps of 1 across rows}\\ \text{here } N = 4\end{array}$$

$(1, 1, 1, 1)$ is a left eigenvector with $\lambda = 1$. The right eigenvector yields the values $\phi(0), \ldots, \phi(N-1)$ at the integers. The recursion determines ϕ at all dyadic points. Values at other points are never used.

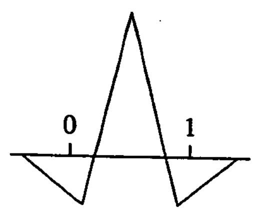
1.2. Wavelets and orthogonality. Finally we define a wavelet. It comes from the scaling function ϕ by taking "differences":

$$W(x) = \sum(-1)^k c_{1-k}\phi(2x - k). \tag{A.6}$$

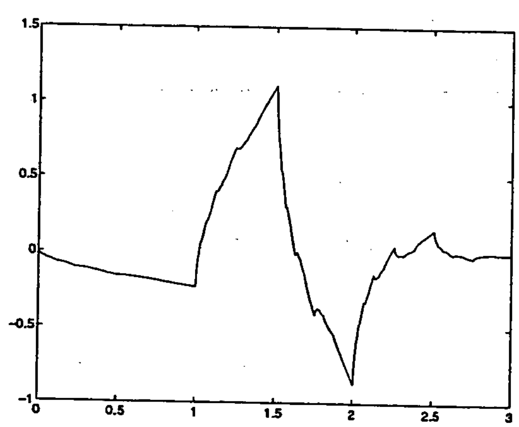
We write W in place of the usual ψ, to distinguish more clearly from ϕ. Notice $2x$ on the right, and especially $(-1)^k$. Examples show the effect of alternating signs:



Haar wavelet from box function
$W_2(x) = \phi(2x) - \phi(2x - 1)$



"Wavelet" from hat function
$W = \phi(2x) - \frac{1}{2}\phi(2x - 1) - \frac{1}{2}\phi(2x + 1)$



$W_4(x)$ from $\phi = D_4$ Orthogonal wavelet

The wavelet from the hat function does not belong here. It is not orthogonal to $W(x+1)$. The point is that the other two do belong. The Haar function is orthogonal to its own translations and dilations. Historically it was the original wavelet (but with $p = 1$ and poor approximation). The orthogonal wavelet W_4 has $p = 2$ and second-order approximation.

Without formulas for D_4 and W_4, how is the orthogonality of their translates known? We need a test that applies to the recursion coefficients c_k, or to the symbol $P(\xi) = \frac{1}{2}\sum c_k e^{ik\xi}$.

Condition O.

$$|P(\xi)|^2 + |P(\xi+\pi)|^2 \equiv 1 \quad \text{or} \quad \sum c_k c_{k-2m} = 2\delta_{0m}.$$

With this condition, the infinite matrix L^*L in Part 2 is an orthogonal projection. To see now the role of Condition O, suppose the functions $\phi_0(2x-k)$ are orthogonal. Then so are the translates of $\phi_1(x) = \sum c_k\phi_0(2x-k)$:

$$\begin{aligned} \int \phi_1(x)\phi_1(x-m)\,dx &= \int \left(\sum c_k\phi_0(2x-k)\right)\left(\sum c_l\phi_0(2x-2m-l)\right)dx \\ &= \sum c_k c_{k-2m}\int \phi_0^2(2x)\,dx = 0 \quad \text{for } m \neq 0. \end{aligned} \tag{A.7}$$

Construction 1 creates ϕ by iteration from the box function, which is orthogonal to its translates. Therefore (as Daubechies observed) so is ϕ.

The wavelet $W(x)$ in (6) is also orthogonal to $\phi(x-m)$. This is simple but neat, not involving Condition O. The sum in (7) changes to

$$\sum (-1)^k c_{1-k}c_{k-2m}, \quad \text{which is identically zero!} \tag{A.8}$$

Just replace k by $1-n+2m$. This identity is $HL^* = 0$ in Part 2. Then (6) makes $W(x)$ orthogonal to $W(2x-m)$. The orthogonality of $W(x)$ and $W(x-m)$ comes back to Condition O.

The goal in constructing wavelets is to satisfy Conditions A and O. The basic family W_2, W_4, $W_6, \ldots$ was discovered by Daubechies, following Haar's W_2. The accuracies are $p = 1, 2, 3, \ldots$ and there are $2, 4, 6, \ldots$ nonzero coefficients c_k. The smoothness also increases with p—but only by about $\frac{1}{2}$ derivative each time. D_4 and W_4 are Hölder continuous with exponent $.550\cdots$. In Galerkin's method for solving differential equations, it is natural for these wavelets to be the trial functions—broader support than splines, nonsymmetric but orthogonal, multigrid built in, all computations based on recursion, difficulty to be expected at boundaries. The first experiments by Glowinski, Lawton, and Ravachol are particularly interesting for Burgers' equation.

2. Algorithms for wavelet expansions

Now comes a change of direction. Instead of discussing the properties of wavelets, we describe algorithms. The main question is how to *decompose* a signal into its wavelet coefficients, and how to *reconstruct* the signal from the coefficients. There is a "tree algorithm" or "pyramid algorithm" that makes these steps simple and fast. It does for the discrete wavelet transform what the Fast Fourier Transform (FFT) does for the discrete Fourier transform. The algorithm is fully recursive.

The user chooses a specific wavelet. We begin with the simplest choice, based on the box function. It satisfies the orthogonality property (Condition O), so all pieces of the decomposition are orthogonal. The approximation property (Condition A which preserves polynomials) determines how quickly the coefficients decay—for efficiency we want to stop the decomposition early. In that respect the box function is poor. Efficiency is the reason for working with

higher wavelets $W_4, W_6, W_8, \ldots$, and simplicity is the reason for starting with W_2. This is Haar's wavelet $[1\ -1]$.

The discussion will be discrete—for vectors not functions. We are given $n = 2^J$ values $f_1, \ldots, f_n$. They may be equally spaced values of a function $f(x)$ on a unit interval. The goal is to split this vector f into its components at different scales, indexed by j. At each new level the meshwidth h is cut in half and the number of wavelet coefficients is doubled. The decomposition is

$$f = f^{\phi} + f^{(0)} + \cdots + f^{(J-1)}.$$

The "detail" $f^{(j)}$ is a combination of 2^j wavelets at scale 2^{-j}, and f^{ϕ} is a multiple of the scaling function ϕ. For a numerical example take $J = 2$. Then the finest detail $f^{(1)}$ is the sum of two terms, here with coefficients $b_{11} = 4$ and $b_{12} = 1$:

$$f = \begin{bmatrix} 9 \\ 1 \\ 2 \\ 0 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} + 4\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}. \tag{A.9}$$

Notice that the four components are mutually orthogonal. There are $1+2+\cdots+2^{J-1}$ wavelet coefficients, and the one from f^{ϕ} makes 2^J.

How are the coefficients 3, 2, 4, 1 computed from f? *On the finest scale first.* As in the FFT, the decomposition begins with a "butterfly":

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{2} & -\frac{1}{2} & & \\ & & \frac{1}{2} & \frac{1}{2} \\ & & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 9 \\ 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 1 \\ 1 \end{bmatrix}. \tag{A.10}$$

This is followed by a permutation, in which high frequencies go to the bottom:

$$\begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 1 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 4 \\ 1 \end{bmatrix}. \tag{A.11}$$

The next step is another butterfly, on low frequencies only:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{2} & -\frac{1}{2} & & \\ & & 1 & \\ & & & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 1 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 4 \\ 1 \end{bmatrix}. \tag{A.12}$$

The result is the set of wavelet coefficients 3, 2, 4, 1. The product of the three matrices in (10–12) is the decomposition matrix D. Its inverse is the reconstruction matrix R:

$$D = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \quad \text{has} \quad D^{-1} = R = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix}.$$

The coefficients 3, 2, 4, 1 enter the vector $b = (b_{\phi}, b_{01}, b_{11}, b_{12})$. The wavelet expansion in (9) is $f = R\,b$. The coefficients are $b = R^{-1}f = Df$. This product Df was computed

recursively, from two butterfly matrices with a permutation between. In general there will be J matrices with permutations between.

The reconstruction is also recursive. It inverts (12) then (11) then (10). The global matrix R is the product of these local inverse matrices.

Notice that the operation count is proportional to n. It is best possible (the FFT count is $n\log_2 n$). There are only $n-1$ individual 2-by-2 matrix multiplications, since high frequency coefficients (here 4 and 1) are settled and not reused. The Walsh functions give a different piecewise constant representation, in which the last two basis vectors are $(1,-1,1,-1)$ and $(1,-1,-1,1)$. In that case 4 and 1 enter another butterfly to produce the Walsh coefficients $\frac{5}{2}$ and $\frac{3}{2}$. The Walsh basis is global. The wavelet basis is local, but scaled—its support has width $O(2^{-J})$ at the finest scale and $O(1)$ at the coarsest scale.

Notice also the normalizing factors $\frac{1}{2}$ in decomposition (and 1's in reconstruction). The alternative is to introduce $1/\sqrt{2}$ for both. This has the advantage of normalizing the wavelets $W_{jk} = 2^{j/2}W(2^jx-k)$ at every scale. The whole basis is orthonormal (when $\|W\| = 1$). In the discrete case R and D become orthogonal matrices:

$$\widehat{D} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \quad \text{has } \widehat{R} = \widehat{D}^{-1} = \text{transpose of } \widehat{D}.$$

Based on the Haar example, we now start on Mallat's beautiful *tree algorithm* for wavelets. The simple average from $\left[\frac{1}{2}\ \frac{1}{2}\right]$ is replaced by a discrete filter based on ϕ. The difference $\left[\frac{1}{2}\ -\frac{1}{2}\right]$ is replaced by a filter based on W. The filters use the same recursion coefficients c_k that led to ϕ and W in the first place.

Decomposition. The given n-vector f is on the finest scale $h = 2^{-J}$. The *fine-to-coarse filter* (the "restriction operator" in multigrid language, the lowpass filter in signal processing language) is L. It produces a vector with half as many entries:

$$(Lf)_i = \frac{1}{2}\sum c_{2i-j}f_j, \quad i = 1,\dots,\frac{n}{2}. \tag{A.13}$$

In the Haar example with $c_0 = c_1 = 1$, the entries of Lf are $\frac{1}{2}(f_1+f_2)$ and $\frac{1}{2}(f_3+f_4)$. The recursion continues to coarser scales, and after J steps it reaches a single number—the coefficient b_ϕ in f^ϕ at the coarsest scale $h = 1$. Here $b_\phi = \frac{1}{4}(f_1+f_2+f_3+f_4)$.

The dual to L is the *coarse-to-fine map* L^* (the "interpolation operator" in multigrid language). Notice the change of index and the disappearance of $\frac{1}{2}$:

$$(L^*g)_j = \sum c_{2i-j}g_i, \quad j = 1,\dots,n. \tag{A.14}$$

In the Haar example L^*Lf has entries $\frac{1}{2}(f_1+f_2), \frac{1}{2}(f_1-f_2), \frac{1}{2}(f_3+f_4), \frac{1}{2}(f_3-f_4)$. It is the projection of f onto the subspace that is piecewise constant at scale $2h$. It gives a blurred picture, with details lost.

The decomposition picks out these details, orthogonal to the average. The projection onto the wavelet subspace is the high frequency component:

$$f^{(J-1)} = f - L^*Lf. \tag{A.15}$$

This repeats at every stage. There is an "average" or "blurred picture" $a^{(j-1)} = La^{(j)}$, starting from $a^{(J)} = f$. The detail lost in that average is the component of f at that stage:

$$f^{(j-1)} = (I - L^*L)a^{(j)} = a^{(j)} - L^*a^{(j-1)}. \tag{A.16}$$

This is a first statement of the decomposition algorithm. We will see how Condition O simplifies the formula.

Reconstruction. To produce f from its details $f^{(j)}$, run the recursion (16) in reverse:

$$a^{(j)} = f^{(j-1)} + L^* a^{(j-1)}. \tag{A.17}$$

This starts from the coarsest detail $f^{(0)}$ and the totally blurred picture $a^{(0)} = f^{\phi}$. It returns to $f = a^{(J)}$.

Apply orthogonality. The most elegant part of the algorithm is still to come. It is not nec-
ssary to compute the detail vector $f^{(j)}$ from (16), and then to compute its wavelet coefficients b_{jk}. Those are the numbers we want (4 and 1 in the example at level $j = 1$). *These numbers an be found directly from* $a^{(j)}$.

Review the Haar example first. The lowpass filter gave $a^{(1)}$ from $f = a^{(2)}$:

$$Lf = \frac{1}{2}\begin{bmatrix} c_1 & c_0 & & \\ & & c_1 & c_0 \end{bmatrix}\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & & \\ & & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 9 \\ 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

he blurred picture is $a^{(1)} = (5, 5, 1, 1)$. At the next level the low-pass filter leaves 3, the co-fficient of $(1, 1, 1, 1)$. *We now want the orthogonal filter—the highpass filter H.* In the Haar xample it produces

$$Hf = \frac{1}{2}\begin{bmatrix} c_0 & -c_1 & & \\ & & c_0 & -c_1 \end{bmatrix}\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & & \\ & & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 9 \\ 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}.$$

hose coefficients 4 and 1 represent the detail $f^{(1)} = (4, -4, 1, -1)$, which is lost when $a^{(2)}$ blurred to $a^{(1)}$. At the next level H is applied to $a^{(1)}$. That produces $\frac{1}{2}(5) - \frac{1}{2}(1) = 2$. This the coefficient b_{01}, representing the detail $(2, 2, -2, -2)$ lost when $a^{(1)}$ is blurred to $a^{(0)}$. We ow put these pieces together into *Mallat's pyramid algorithm*:

ecomposition. Initialize $a^J = f$. For $j = J, \ldots, 1$ compute

$$a^{j-1} = La^j \quad \text{and} \quad b^{j-1} = Ha^j. \tag{A.18}$$

econstruction. Start with a^0 and $b^0, \ldots, b^{J-1}$. For $j = 1, \ldots, J$ compute

$$a^j = L^* a^{j-1} + H^* b^{j-1}. \tag{A.19}$$

ie full decomposition is represented by a tree of filters:

$$\begin{tikzcd} a^J \arrow[r, "L"] \arrow[dr, "H"'] & a^{J-1} \arrow[r, "L"] \arrow[dr, "H"'] & a^{J-2} & \cdots \arrow[r, "L"] \arrow[dr, "H"'] & a^0 \\ & b^{J-1} & b^{J-2} & & b^0 \end{tikzcd}$$

ie reconstruction goes from the branches of the tree back to the root:

$$\begin{tikzcd} a^0 \arrow[r, "L^*"] & a^1 \arrow[r, "L^*"] & a^2 & \cdots \arrow[r, "L^*"] & a^J = f \\ b^0 \arrow[ur, "H^*"'] & b^1 \arrow[ur, "H^*"'] & & \arrow[ur, "H^*"'] & \end{tikzcd}$$

The next step is to identify these filter matrices L and H for examples other than "box and Haar."

Note. The filter matrices L and H have half as many rows as columns. By dropping the parentheses around j, we distinguish the vector a^j with only 2^j components from the vector $a^{(j)}$ with the full $w^J = n$ components. The vector a^j contains the expansion coefficients of $a^{(j)}$ with respect to the translates $\phi(2^j x - k)$. See the example above and the multiresolution below!

2.1. The filter matrices L and H. The matrix L is known from the first part of the paper. Its entries $L_{ij} = c_{2i-j}$ are the recursion coefficients for the scaling function. Rows 1, 2 and columns $-1, 0, 1, 2$ are displayed with $N = 3$:

$$L = \frac{1}{2}\begin{bmatrix} c_3 & c_2 & c_1 & c_0 & & \\ & & c_3 & c_2 & c_1 & c_0 \end{bmatrix}.$$

The beautiful thing is that the highpass filter (strictly speaking it is band-pass) uses the same coefficients. H is associated with the wavelet W just as L is associated with the scaling function ϕ. Equation (6) for W uses the same c_k, but with alternating signs and reversed order. The wavelet filter has

$$H_{ij} = (-1)^{j+1} c_{j+1-2i}. \tag{A.20}$$

Rows 1, 2 and columns 1, 2, 3, 4 are displayed:

$$H = \frac{1}{2}\begin{bmatrix} c_0 & -c_1 & c_2 & -c_3 & & \\ & & c_0 & -c_1 & c_2 & -c_3 \end{bmatrix}.$$

The indices were chosen to match the Haar example (variants are possible). The transposed matrices, without the factor $\frac{1}{2}$, represent the dual filters L^* and H^*. The important points now come quickly, and matrix multiplication is the best proof.

Theorem 1. *By their construction the filters are orthogonal:*

$$HL^* = 0. \tag{A.21}$$

This multiplication is the reason behind the construction of H — alternating signs, reversed order, index shifted by one. See equation (8).

We finally come to the reward for Condition O: $\sum c_k c_{k+2m} = 2\delta_{0m}$. The reason for that condition is in the reward. Remember that the box function and D_4 satisfied this requirement but not the hat function or the cubic spline. Condition O can be stated and understood in transform space, but I believe that the matrix interpretation is again the clearest.

Theorem 2. *If condition* O *holds then*

1. $LL^* = I$ *and* $HH^* = I$. (A.22)
2. L^*L *and* H^*H *are mutually orthogonal projections with*

$$L^*L + H^*H = I. \tag{A.23}$$

Remember that L and H map into subspaces half as large as the original. L^* and H^* map back. The identity operators in (22) are on the half-sized subspaces.

The proof of (22) is by direct matrix manipulation. Condition O gives the result. Then it follows that $L^*LL^*L = L^*L$, so L^*L is a projection — and similarly for H^*H. The property $HL^* = 0$ in (21) yields $H(L^*L + H^*H) = H$. The transpose $LH^* = 0$ yields $L(L^*L + H^*H) = L$. The operator in (23) is the identity on both orthogonal components — the ranges of L and H — so it is the identity. We have an orthogonal decomposition by "*quadrature mirror filters*" L and H at every step.

2.2. Multiresolution of L^2. The last paragraphs changed quietly from functions to vectors. That was for the sake of algorithms, which use values of ϕ and W at dyadic points $k/2^j$. The Haar example began with f at equally spaced points on (0, 1]. But the filter matrices really apply to discrete values along the whole line — they are infinite matrices. More than that, the decomposition $f = \sum f^{(j)}$ is just as valuable for functions in L^2 as for vectors in l^2.

This multiresolution yields the details of f at all scalings 2^{-j}. On the whole line we take $j = 0, \pm 1, \pm 2, \ldots$. The decomposition develops an idea that was already present in approximation theory — to put frequencies together in "octaves." (Besov spaces combine frequencies $2^j \leq \xi < 2^{j+1}$. It seems that the ear also receives frequencies on a logarithmic scale.) For functional analysis the starting point is the subspace S_j spanned by the translates $\phi(2^j x - k)$. If a function $g(x)$ is in S_j, then $g(2x)$ is in S_{j+1}. The dilation equation writes $\phi(x)$ as a combination of $\phi(2x - k)$, which assures that $S_0 \subset S_1$. At all scales we have

$$\cdots S_{-1} \subset S_0 \subset S_1 \subset S_2 \cdots \text{ with } \cup S_j \text{ dense in } L^2 \text{ and } \cap S_j = \{0\}.$$

Now turn to the *wavelet subspace* W_j. It is spanned by the translates $W(2^j x - k)$. It is invariant under translation by multiples of 2^{-j}. If $g(x)$ is in W_j then $g(2x)$ is in W_{j+1}. The construction $W(x) = \sum(-1)^k x_{1-k}\phi(2x-k)$ puts W and its translates into S_1, and makes them orthogonal to S_0. *In fact, W_0 and S_0 are orthogonal complements in S_1*. At every scale $W_j \bigoplus S_j = S_{j+1}$. *The spaces S_j give the "partial sums" of the differences W_j:*

$$\cdots \bigoplus W_{-1} \bigoplus W_0 \bigoplus \cdots \bigoplus W_j = S_{j+1} \quad \text{and} \quad \bigoplus_{-\infty}^{\infty} W_j = L^2.$$

The multiresolution of f is a splitting into components $f^{(j)} \in W_j$:

$$f = \sum_{-\infty}^{\infty} f^{(j)} \quad \text{or} \quad f = f^{\phi} + \sum_{0}^{\infty} f^{(j)}, \quad f^{\phi} \in S_0. \tag{A.24}$$

This is a very satisfying decomposition of L^2 functions, classical but with new subspaces. The coefficients b^j in Mallat's pyramid algorithm corresponded to $f^{(j)} \in W_j$, and a^j corresponded to $a^{(j)} \in S_j$.

The analogue of the discrete Fourier transform was in the algorithm. The analogue of ordinary Fourier series is (24). The analogue of the Fourier integral formula is the *integral wavelet transform*. Representations of different groups give rise to different transforms.

2.3. Applications. Image processing works with $F(x, y)$, so it is natural to look for *two-dimensional wavelets*. The simplest construction uses the products $\phi(x)\phi(y)$, $\phi(x)W(y)$, $W(x)\phi(y)$, $W(x)W(y)$. Orthogonality is clear. New constructions have been invented that are genuinely two-dimensional, but it is useful to start with the tensor products of "box and Haar." The given two-dimensional array F yields a two-dimensional array B of wavelet coefficients.

For pattern recognition, a major difficulty with the wavelet transform B is the lack of translation invariance. If the pattern is shifted by a fraction of h, its wavelet model is changed. A higher sampling rate is possible but expensive. Mallat studies instead the *zero-crossings* of the wavelet transform, which locate the signal edges. Now the difficulty is to make the reconstruction stable. In edge detection the first wavelets were Laplacians of shifted Gaussians, introduced by Gabor. The orthogonal wavelets of Meyer are C^∞ with polynomial decay, the Battle-Lemarié wavelets based on splines are C^n with exponential decay, and the Daubechies wavelets are C^n (smaller n) with compact support.

In closing we recall the original problem — to localize in time and frequency. Geophysics needs to represent short high-frequency pulses. Physics needs to divide up phase space. The coherent states $g_{pq} = e^{ipx}g(x-q)$ give a "Weyl-Heisenberg" frame, with some redundancy — but still f can be reconstructed from $\iint (f, g_{pq})g_{pq}\,dp\,dq$. Mathematics needs (or wants) an orthogonal decomposition, better than g_{pq} at high frequencies and with no redundancy. The answer for now is wavelets.

It is a pleasure to thank Ingrid Daubechies and Howard Resnikoff for introducing me to wavelets.